HYBRID IMAGE ANNOTATION IN FOLKSONOMIES USING TAGS/WORDS CO-OCCURRENCES

Meshesha Legesse Bedane

A Dissertation Submitted to IT Doctoral Program

Presented in Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Information Technology (Information Retrieval)

Addis Ababa University
Addis Ababa, Ethiopia

October 2018

Addis Ababa University

School of Graduate Studies. IT Doctoral Program

This is to certify that the thesis prepared by Meshesha Legesse Bedane, entitled: "**Hybrid Image Annotation in Folksonomies Using Tags/Words Co-occurrences**" and submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy in Information Technology (Information Retrieval) complies with the regularities of the University and meets the accepted standards with respect to originality and quality

Signed by Examining Committee:

Abebe Rorissa (PhD) External Examiner	Abebe Rorissa Signature	29 Oct 2018 Date
Wondwossen Mulugeta (PhD) Internal Examiner	Signature	09 Oct/2018 Date
Gabriele Gianini (PhD) Advisor	Signature	03 Oct 2018 Date
Dereje Teferi (PhD) Co-advisor	Signature	03/04/2018 Date

Dr Dereje Teferi

Track Coordinator or IT PhD Program Director

Addis Ababa University

School of Graduate Studies, IT Doctoral Program

This is to certify that the thesis prepared by the candidate, entitled: "Hybrid Image Annotation in Folksonomies Using Tags/Words Co-occurrences" and submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy (Information Retrieval) complies with the regularities of the University and meets the accepted standards with respect to originality and quality.

Signed by Examining Committee:

Abebe Rorisaa(PhD)

External Examiner	Signature	Date

Wondwossen Mulugeta(PhD)

Internal Examiner	Signature	Date

Gabriele Gianini(PhD)

Advisor	Signature	Date

Dereje Teferi(PhD)

Co-advisor	Signature	Date

Track Co-coordinator or Director of IT PhD Program

I would like to dedicate this thesis to my **Mother Emahoy Tewabech Gezmu Abayire and My Father Legesse Bedane Cheru**

Declaration

I, hereby, declare that except where specific reference is made to the work of others, the contents of this dissertation are based on my research works and have not been submitted for consideration for any other degree or qualification in this, or any other university.

This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgments.

Meshesha Legesse Bedane

October 2018

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Prof. Gabriele Gianini for his unreserved support during my Ph.D study, for his patience, motivation, enthusiasm, and immense knowledge. His guidance, follow up and encouragement helped me in all the time of research and writing of this thesis. Beyond his professional support, he helped me in facilitating research trips to his University(Milan University) and INSA at Lyon and finding funds for publication related fees. Thank you so so much!

I would like to express my heartfelt appreciation to my co-advisor Dr. Dereje Teferi. Whenever I faced difficulties, he was always there to provide me his invaluable advices and supports. He did an unlimited effort to prepare me in writing of this thesis. He took his much time for reading previous drafts of this dissertation and providing many valuable comments that improved this study.

I would like, also, to thank the rest of my thesis committee : Dr. Solomon Atnafu and Dr. Fekade Getahun, for their encouragement and insightful suggestions and comments.

I would like to extend my gratitude to IT Doctoral Program members for their all rounded unlimited support.

Last but not the least, I would like to thank my family: my brother Bassaznew Legesse Bedane and his wife Addisalem Abebe, my nephew Gebeyehu Yigezu, and my lovely wife Beza Assfa Demissie, for their love and support. I owe them lots of time.

Abstract

With the development of Web 2.0 and the launch of Web Sites like *Flickr*, *sharing and collaboratively annotating images(tagging in Folksonomies)* with keywords, called *tags*, are becoming very popular. Although tagging simplifies resource browsing and retrieval, it suffers from several issues. Among the issues are *redundancy* and *ambiguity*. Sometimes a tag which is a very important element will be missed, if the user uploads image without tag.This thesis proposed a hybrid image annotation technique which consists of both user assisted(semi-automatic) and automatic image annotation strategies. The study mainly focuses on the problem of (1) resolving tag word-sense ambiguity(tag-word disambiguation) within a typical semi-automatic tagging procedure, and (2) Recommending tags of the new image automatically, if it is uploaded without tags using tags of previously uploaded similar images based on the result of tags (or words) co-occurrence analysis.

Both should rely on effective word-to-context relatedness metrics. Among the most effective relatedness metrics are those defined on the basis of a feature vector representation of the words. In the study comparison of different word-to-context relatedness metrics in terms of effectiveness within finding tags (or words) relatedness process is done. Based on the results of the comparison, we propose to use a metrics derived from a Maximum Likelihood estimator of the Jensen-Shannon Divergence among feature-count histograms and we show that such a metrics performs(in terms of quality of the output) better than both the Jensen-Shannon and the Symmetrized Kullback-Leibler divergence between histograms. The relative gain in quality within the task of unsupervised cue-word discovery and tags co-occurrence analysis by using a synthetic language corpus has been studied.

In tags relatedness analysis using co-occurrence information, a word is assigned to a specific context chosen among the different ones to which it is related. Relatedness to a context is often defined based on the co-occurrence of the target word with other words (context words) in sentences of a specific corpus. Context words play the role of features for the target. The overall disambiguation process or tags co-occurrence analysis can be thought as a classification process. A problem with this approach is that a large number of possible context words can reduce the classification performance, both in terms of computational effort and in terms of quality of the outcome. Feature selection can improve the process in both regards, by reducing the overall feature space to a manageable size with high information content. In this work, in disambiguation or tags co-occurrence analysis, a novel approach using a feature selection based on the *Shapley Value (SV)* – a Coalitional Game Theory related metrics, measuring the importance of a component within a coalition is proposed. By including in the feature set only the words with highest Shapley Value, tags quality(correctness of tags) and performance improvements are obtained. The problem of the exponential complexity in the exact SV computation is avoided by an approximate computation based on sampling. The study demonstrates the effectiveness of this method and of the sampling approach results, by using both a synthetic language corpus, a corpus prepared from Flickr images database(previously annotated images) and a real world linguistic corpus from Wikipedia English document dump.

We showed the extent to which each of the procedures in our approach contributes to the overall performance improvements using standard evaluation metrics.

keywords: *Tagging, disambiguation, semantic relatedness, dissimilarity metrics, Jensen-Shannon divergence, Flickr, Folksonomy, Feature selection, Shapley Value, dimensional reduction.*

Table of contents

List of figures

List of tables

List of Acronyms

AJSD	Adapted Jensen–Shannon Divergence
API	Application Programming Interface
CBIR	Content Based Image Retrieval
CCA	Canonical correlation analysis
CMRM	Cross-Media Relevance Model
CRM	Continuous-space Relevance Model
EM	Expectation Maximization
FOWs	Frequently Occurring Words
FS	Feature Selection
GPS	Global Positioning System
JSD	Jensen–Shannon Divergence
KLD	Kullback–Leibler Divergence
LDA	Latent Dirichlet Allocation
LS	Laplacian Score
MBRM	Multiple Bernoulli Relevance Model
ML	Maximum Likelihood
PCA	Principal Component Analysis
SIFT	Scale-Invariant Feature Translation
Sh	Shapley
SVD	Singular Value Decomposition
TBIR	Text Based Image Retrieval

TPC	Probabilistic Topic-Connection
VLM	Visual Language Model

Chapter 1 Introduction

1.1 Background of the Research

With the development of Web 2.0 (evolution from web 1.0 to web 2.0) and the launching of web sites for social media(like del.icio.us[1], Flickr[2], CiteLike[3], Youtube[4]), social tagging (or collaborative tagging or social classification, social indexing or folksonomy) is becoming more and more popular (Cantador et al., 2011; Dattolo et al., 2010; Dellschaft and Staab, 2008; Deshmane and Wankhade, 2014; Doerfel et al., 2016; Thielen, 2010; Wang et al., 2013).

With these developments, in a way that is easy-to-use, interactive, and participatory scenarios, content creation and sharing features have become quite popular recently in the World Wide Web by highly active involvements of users(Benz et al., 2008). Users contribute their contents (e.g., bookmarks, pictures, audios, videos and the like) by adding their opinions in the form of labels (or by annotating content). This means that, the Web 2.0 now offers a platform for new means of interactions and has shifted more power and influence to users. The way of using the Internet is changing in such a way that every Internet user is not only a *consumer* but also a *producer* (Mousselly-Sergieh et al., 2013; Quintarelli, 2005).

In social tagging systems, users are organizing, sharing and retrieving digitized resources by enhancing their quality with semantically invaluable informa-

[1] https://del.icio.us/
[2] https://www.flickr.com
[3] https://www.citeulike.org/
[4] https://www.youtube.com/

tion (or keys) (Cattuto et al., 2006; Fu et al., 2010; Gemmell et al., 2008a; Kim et al., 2010; Noll and Meinel, 2008; Subramanya and Liu, 2008). For instance, in Delicious-users assign labels to URLs, whereas in Flickr to photos uploaded by them or by others (Gupta et al., 2010; Quintarelli, 2005).

Flickr, the most popular social tagging system for images (Ivanov et al., 2010), has facilities which encourage users to supply labels, which are crucial contents for the organization and communication of images, in the form of freely chosen texts for the images they have been uploading (Gemmell et al., 2008a; Yan et al., 2007) and the majority of labels in this sites are created by the users who have their photos themselves (Mathes, 2004a). These freely chosen texts are called tags (Barclay, 2009; Lee et al., 2010; Lipczak and Milios, 2010; Milicevic et al., 2010; Weinberger, 2005; Yin et al., 2010) and these tags are used to bridge *the semantic gap*, hence they have direct advantage for users for a number of purposes such as browsing, retrieving and categorization (Bischoff et al., 2008). *semantic gap* is defined by (Smeulders et al., 2000) as follows:

> *"... the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation".*

Because of the facilities provided by the Web, users have been encouraged to upload images on the Web for sharing with others. Flickr can be a typical example. The trend shows the number of photos uploaded to this site is increasing (Figure 1.1[5]). The vast amount of images available on the Web demanded to

[5] https://www.flickr.com/photos/franckmichel/6855169886

have a technique which will help users to search relevant images of their interest effectively and efficiently, and this can be achieved through annotations, so that users can use keyword(s)-based query which tries to match tags or annotations on images so that images whose tags or annotations are highly relevant to the query are returned (Chen et al., 2012; Sigurbjörnsson and Van Zwol, 2008).

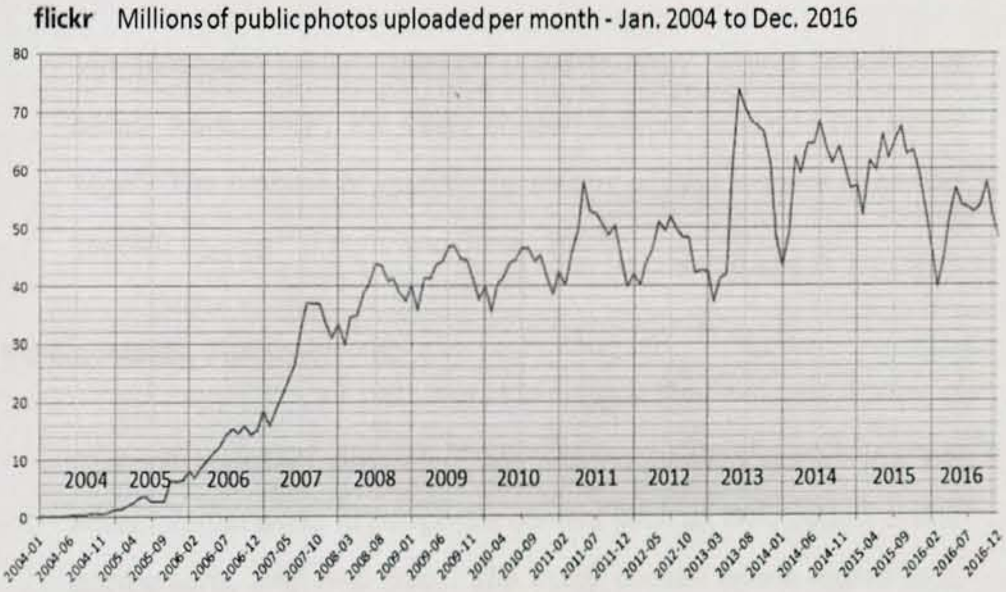

Fig. 1.1 Number of photos uploaded to Flickr from 2004 to 2016

Although the technological developments in recent years are improving in bridging the "semantic gap" - the difference between low-level visual features and high-level concept semantics (Smeulders et al., 2000; Wang et al., 2012), by attaching high level tags to images, challenges in this area are still attracting researchers from different communities. As a consequence, a vast number of image annotation techniques have been proposed. One such techniques is assigning tags to images, which can increase their semantics- the association between low-level visual features and high-level concepts that can be described in words, through a collaborative way on a *collaborative tagging systems (or Folksonomies)* (Lindstaedt et al., 2009; Seneviratne and Izquierdo, 2011; Smith, 2004) and these systems are progressing (Vander Wal, 2007).

The terms Folksonomy, a term first coined by Thomas Vander (Vander Wal, 2007), Tags and Social Tagging are described as follows:

> *"An annotation system open for users to apply subject headings is called "folksonomy", the freely chosen subject headings are called "tags". The process of indexing by means of folksonomies is named "(social) tagging"" (Peters and Stock, 2007).*

More generally, *Folksonomy* is:

> *"...a collaboratively generated, open-ended labeling system that enables Internet users to categorize content such as Web pages, online photographs, and Web links. The freely chosen labels, called tags, help to improve the effectiveness of search engines because content is categorized using a familiar, accessible, and shared vocabulary. The labeling process is called tagging" (Curtis et al., 2012).*

In this description there are three things: Users, Items(or Documents) to tag and Tags as depicted in Figure 1.2. Most research works are modeling folksonomy basically as s set of triplets, user-tag-resource, to mean that a given user supplied a particular tag to a certain resource (Mika, 2005; Wu et al., 2006). Folksonomy systems relate resources through tags as well as through the users who provide tags to the resources. In social tagging, group of people who are annotating resources with freely selected texts(tags) is presented by the authors in (Milicevic et al., 2010) is shown in the figure 1.2.

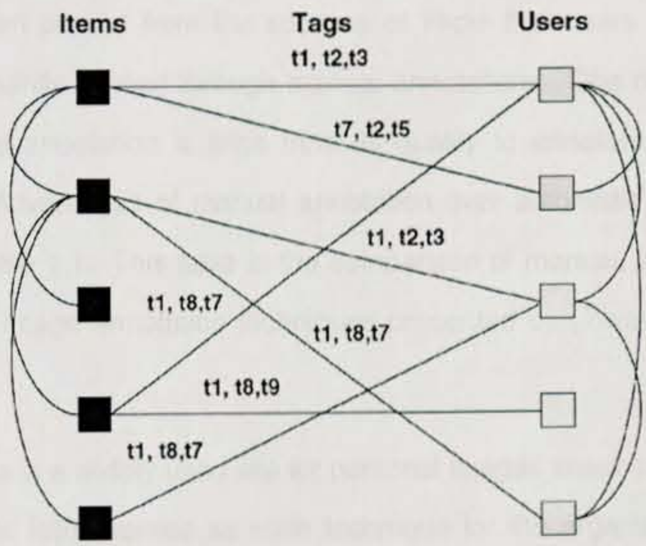

Fig. 1.2 Items, tags and users in a folksonomy.

Folksonomy formally described by the authors in (Hotho et al., 2006b) as a Tuple F:= (U, T, R, Y) where U, T, and R are finite sets, whose elements are users, tags, resources and Y is a ternary relation between them. where T is the set of tags contributed by a set of users U to annotate a set of resources R. Y co-occurrence of two tags $t_1;t_2 \in T$ indicates that they are used by one or more user to describe the same resource $r \in R$.

In Folksonomies, users or social inputs with their added context represent a strong substitute for expert annotations (Sawant et al., 2011). Users' collaboration in tagging is to provide tags for different purposes-for future retrieval, contribution, and personal references.

There is an increased interest of using manual image annotation systems for organizing and communication purpose of multimedia content (Min et al., 2009).This interest has been proved from the success of Flickr that users are willing to provide the semantic context through manual annotations. One of the reasons of using manual annotation is arise from its quality to annotate images (Yan et al., 2007). Advantages of manual annotation over automatic annotation is presented in Table 1.1. This table is the comparison of manual, automatic and semi-automatic image annotation techniques presented by (Tiwari and Kamde, 2015).

Flickr website is a widely used site for personal images sharing among users mainly based on folksonomies as main technique for the organization of their images. Flickr images database is alarmingly increasing in number of images, users, and groups.This rapid growth is because of the development of the Internet and the ease of accessibility of image capturing digital devices (Liu et al., 2007b).

In Flickr images database, every time new images are added, they have to be annotated (or tagged) for future retrieval (Lindstaedt et al., 2009). Users naturally assign tag to some of shared images and leaving others without tags, or with incomplete metadata. This inconsistency of metadata/information extremely deteriorates search, since images without appropriate tags are much difficult to retrieve (Ivanov et al., 2010).

Tagging in Folksonomy has become a technique for labelling and organizing digital contents on the web. Although tagging in these systems is simplifying the organization and sharing of these web resources, it faces a number of problems. Mathes (Mathes, 2004b) points two main issues of user-supplied tags: *ambiguity*

Table 1.1 Comparison of manual, semi-automatic, and full automatic Image Annotation Techniques

Technique	Advantages	Disadvantages
Manual	Reliable and accuracy in extracting semantic information at several levels	Tedious, requires a lot of time and efforts, costly, highly subjective.(The perspective of tags or textual information given by an annotator can be different from the perspective of a user. A picture can mean different things to different people and to the same person at different times)
Semi-Automatic	Better than manual annotation in terms of efficiency more accurate than automatic annotation (It combines the efficiency of automatic annotation and the accuracy of manual annotation of images)and useful for dynamic databases.	Requires user interfaces refinements to improve the feedback process. It depends on the user's interaction to some degree
Automatic	Speed (saves time)	Less reliable than manual, more error- prone, produces more general (less detailed) annotation as compared to manual method, Less accurate than semi- automatic annotation

and lack of synonym control, which is also known as *redundancy* (Gemmell et al., 2009; Rahuma, 2013).

Tag ambiguity arises when the same tag is used to indicate different meanings. Typical examples are word-sense ambiguity (e.g. the word "palm" in different context), language ambiguity (e.g. "Gift" means *poison* in German and *present* in English) (for further details refer to (Weinberger et al., 2008)) and scarcity (or lack of specificity). Tag scarcity is when a certain tag is insufficient to give the required meaning for latter retrieval. For instance, consider a user is tagging an image by the tag "Lalibela" on social tagging system. As we can observe this word is scarce to carry enough information, since we can have more images by this tag (Lalibela City, Lalibela King, Lalibela Church, Lalibela Hotel,...). But if it is tagged as "Lalibela Church", it will have the required meaning sufficiently. Hence the need of developing a technique which will include additional information when an image is tagged by such types of tags accurately and relevantly is of paramount important (Rahuma, 2013). In our proposal, when the user provides the tag "Lalibela", set of cue-words like "Church", "King", "City", "Hotel",.. will be recommended so as the user can select one which will fit to the context in mind. In our case we call all of them tag ambiguities. In this study, we focus on the problem of resolving tag ambiguity within a typical semi-automatic tagging procedure.

Besides, in this tagging, users may not provide tag to the image they are uploading because of different reasons-may not know the content or may be in hurry to tag a given content.

To alleviate these types of problems a *hybrid image annotation technique*, which consists of both user assisted(semi-automatic) and automatic image annotation strategies, is proposed in this study.

1.2 Research Motivation and Statement of the Problem

Because of their easiness to organize and retrieve digital content using freely chosen texts (or tags), the popularity of social tagging systems in social medias like Flickr and Delicious have been boosting (Milicevic et al., 2010).

As a consequence of these, social tagging systems are becoming popular and allowing users to upload and tag resources. Users with different interest and level of understanding can annotate their resource using freely chosen and uncontrolled vocabulary of tags. In this context a tag which is a very important element for the organization and future retrieval purpose can face the problem of ambiguity (Lindstaedt et al., 2009). Ambiguous in a sense that, the tag which has been provided for a given resource may not have same meaning for different users or it can be insufficient to carry the required information. For instance the tag "bank" is ambiguous, since it has many meanings. One can be a financial institution and the other can be side of a river or a sea or an ocean. On the other hand a tag "mountain" for a certain mountain may not be sufficient to provide enough information for a certain user but if it is "mountain dashen" , it will be sufficiently enough, since this is a particular mountain with the name "dashen".

Moreover in this tagging system, the users may not provide tag to the resource they are uploading (Lux et al., 2010; Zhang et al., 2015). For instance, the user might not know the exact tag or name of the resource or they may be in a hurry to provide tag for that resource. The general goal of tagging is to improve the quality of tags so that the retrieval system using these tags will be able to retrieve images (or photos) from large images collections efficiently and effectively. So in line of this goal a hybrid image annotation technique is proposed for this type of tagging system or Folksonomy. In particular,this research focuses on the following issues:

- Resolving tag ambiguity in a semi-automatic image tagging process(Context-based tags recommendation) without processing the content of the image. Context-based is to mean based on the preceding or following part of a text (a word or a tag) to clarify the meaning.
- Tagging images using co-occurrences of tags of previously annotated images from Flickr database automatically(Automatic Image's Tags recommendation).

Based on these issues, this study answers the following research questions:

1. How can the accuracy of relatedness of tags/words be improved to determine cue-words in tag-sense disambiguation process for user provided tag(s) on social tagging systems?
2. How can co-occurrences of user contributed tags on social tagging system (e.g.; Flickr) be exploited to generate tags of new images which are uploaded without tags?

3. How can important tags/words selection can be improved to boost the accuracy of finding words/tags relatedness procedure?
4. How can semantically-enriched content like Wikipedia be exploited or used to find tags relatedness?

1.3 Objectives of the Research

1.3.1 General Objective

The general objective of this work is designing an image annotation mechanism, by analyzing tags'/keywords' relatedness based on co-occurrences information, in a typical social tagging system(or Folksonomy).

1.3.2 Specific Objectives

In order to achieve the general objective, the following specific objectives are formulated:

1. review existing image annotation approaches in order to identify the methods and techniques used so far in both user assisted and automatic image annotations.
2. designing a framework for image annotation in Folksonomy which incorporates both Semi-automatic and Automatic Image tags assignments

3. develop an algorithm which uses the words relatedness information to resolve tags ambiguity and improve tags semantics by exploiting tags relatedness information based on co-occurrence statistics in a typical semi-automatic image tags recommendation procedure.
4. develop a procedure which uses words' importance measure toward the disambiguation algorithm(this is to reduce dimension and to select best annotation words)
5. generate tags of a new image automatically using previously tagged images(this is when a user upload an image without tag)
6. perform experimentation to test the proposed algorithms.
7. evaluate the performance of the proposed approaches using standard evaluation measures(metrics).

1.4 Significance of the Research

The primary area of interest in this research is *Multimedia Annotation* specifically *Image Annotation*. The basic idea is conducting a research on techniques in annotating images in *Folksonomy*. Both text documents and the tags of previously annotated images are used.

The output of the proposed approach of this research can be used by integrating with the available content based image retrieval techniques in social tagging systems. The proposed approach can be a contribution to the industry in resource tagging systems. For example, a given social tagging system can integrate the result into its tagging environment so that a user can upload a picture and use the annotation option to annotate it; the system can help to facilitate images retrieval systems to retrieve images using textual information as search query.

1.5 Scope of the Research

This work focuses on social tagging systems (or Folksonomy) to tag images by giving focus on important element of this technique: the tags set. It uses the available and appropriate content-based image retrieval system to find similar images from the Flickr image database to the target image.The work concentrates on exploiting words/tags relatedness using co-occurrence statistics to propose annotation for the new image which is uploaded with or without initial tag.

1.6 Research Methodology

In this section methods and tools used in this research to achieve the aforementioned specific objectives are presented.

Generally this research work follows the design science methodology for information technology (Peffers et al., 2006; Von Alan et al., 2004) which incorporates the artifacts of information systems developments i.e.; an outcome based information technology research methodology, which offers specific guidelines for evaluation and iteration within research projects.This research paradigm is a problem-solving process and common in information systems (Bichler, 2006;

Hevner and Chatterjee, 2010; Iivari, 2007; March and Smith, 1995) The general process of design science methodology is depicted in figure 1.3 as proposed by (Peffers et al., 2006)

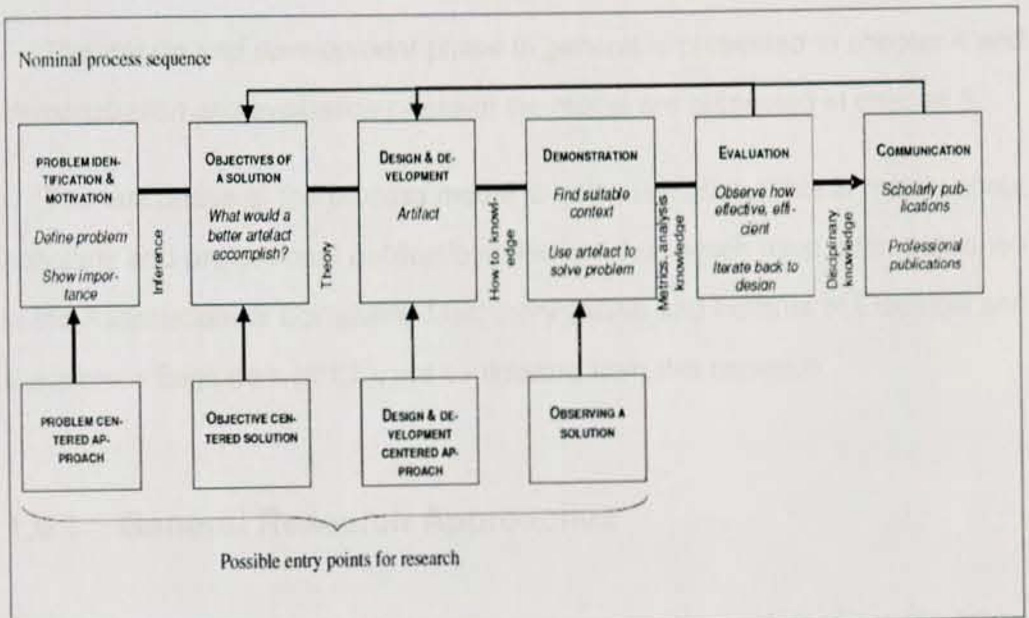

Fig. 1.3 Design science research process model

The first step in the design science research process model is *problem identification*. We have described the motivation, problem identification and research questions in section 1.2 based on the gaps obtained from the background and literature review presented in chapter 2. The starting point of the research is review of existing image annotation techniques in general and social tagging(or tagging in Folksonomy) in particular. From the review, types and status of similarity(dissimilarity) metrics and the approaches to minimize dimensionality and to select best features are explored.

Objective of a solution-is the second phase in the model and we have stated the general and specific research objectives in section 1.3 and proposed approaches to achieve each of the objectives given in section 1.6.1.

The *design and development* phase in general is presented in chapter 4 and *demonstration and evaluation phase* of the model are presented in chapter 5.

The last phase of the process model is *communication*. This is mainly about scholarly and professional publications.Two articles, which have been published in the Association for Computing Machinery (ACM) and Institute of Electrical and Electronics Engineers (IEEE), are contributed from this research.

1.6.1 General Research Approaches

With respect to the individual objectives, the approaches adopted are the followings:

1. review existing image annotation approaches in order to identify the methods and techniques to be used.
   - Investigation of the statistical techniques currently used in semi-automatic and automatic annotations, with specific attention to those which can be considered context-aware. This is to strengthen the awareness and to investigate techniques which will lead to the suggested solutions.
2. propose the general framework of the annotation process

- This is to develop the map of the work in general and to demonstrate what will be included at each stage of the annotation process

3. develop a semi-automatic image annotation procedure

   - This is to develop a procedure which assists the user when they provided a tag for the image they are uploading.
   - The tag would be checked for ambiguity and cue-words would be generated using tags relatedness information and recommended to the user. The user can select the one in their mind and this(the selected one) will be augmented to the original tag(the one which is provided by the user) and assigned as final annotation of the the new image

4. generate tags of a new image automatically using previously tagged images (this is when a user uploads an image without tag)

   - The user may not provide tag to the image they are uploading.Thus the system will generate tags automatically.
   - In this case, using Content-based image retrieval, similar images to the target image are searched and tags of these images are analyzed to assign tags for the new image

A ranking algorithm, to select appropriate annotation words or descriptions to untagged image by analyzing tags (or words) relatedness statistics from tags of previously tagged images and Wikipedia text Corpus, have been used using Shapley Value Analysis

5. develop a prototype to implement the proposed algorithms (development phase of the suggested approaches) and evaluate the performance of the system using standard information retrieval measures.

1.6.2 Datasets and Evaluation Techniques

In this research both synthetic and real word datasets are used. The synthetic dataset(as controlled vocabulary) is used to check the preciseness of the metrics used in the proposed procedure, since real word repositories, despite their size, have limits: typically, they are biased towards their scope and fields. Images are collected from Flickr database with the associated user contributed tags. To use the advantages of semantically-enriched content, text documents from Wikipedia are used. Additionally, images which are annotated by professionals(trained personnel) are used as ground truth.

The standard image annotation systems evaluation techniques are used to assess the performance of our procedures. Most performance measures in these systems are Precision, Recall and F1-scores.

1.7 Thesis Structure

The rest of this thesis is organized as follows:

Chapter 2 provides the detailed literature on the state-of-the-art of image annotation. Particularly brief description on image retrieval and annotation types, and a detailed review on previous works on manual, semi -automatic, automatic and social tagging techniques are presented.

Chapter 3 presents the review work on research efforts closely related to our study, including word relatedness metrics, unsupervised tag senses disambiguation and feature selection techniques.

The general framework, tags relatedness statistics calculations, procedures to resolve tags ambiguities, to select feature tags/word set and to generate tags from previously annotated image are presented in Chapter 4.

Chapter 5 provides the experimentation process, the results obtained and the evaluation which is conducted in order to evaluate the performance of the proposed image annotation approach and strategies on how datasets are prepared and selected.

Finally, chapter 6 presents the conclusion of our research, a summary of tasks achieved by this research, the scope and limitation of the research, and future research areas resulting from this research.

Chapter 2 Literature Review

In this chapter literatures on the state-of-the-art of image annotation in general : such as review work on types of image retrieval, manual, semi and fully automatic image annotations,and collaborative image tagging are presented. Research works on the development of image annotation techniques, usage of tags co-occurrence to improve semantics, and measurements used in each are reviewed and presented. Final, Known probability distribution divergences measures are discussed.

2.1 Image Retrievals

Getting the required data from the web database can be described as retrieval and it can be content or text based, indicated in Figure 2.1 by the work in (Jayaswal and Shrivastava, 2015).

In particular, Image retrieval is the process of searching and extracting images from a large database of digital images (Kulkarni et al., 2008) . Even if they have evolved separately information retrieval, computer graphics, database management and user behavior are interrelated and provide invaluable contribution to the research of Image retrieval which is active since the 1970s.

Image retrieval systems can be Content-based-Figure 2.2,or Text-based-Figure 2.3 presented by (Pal et al., 2013) , or Hybrid based.

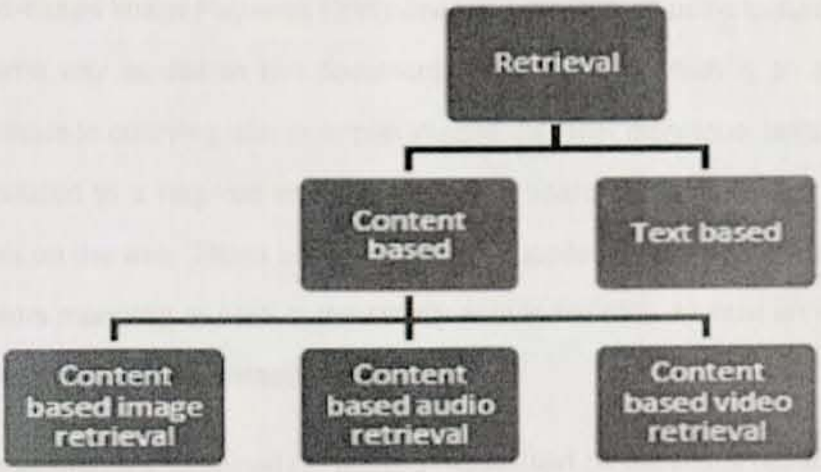

Fig. 2.1 Retrieval Types

In Content Based Image Retrieval(CBIR), the search uses the low level features of the image using image processing techniques (Bulo et al., 2011; Dobrescu et al., 2010; Li et al., 2011; Su et al., 2011; Sumathi et al., 2011). Detailed surveys are found in (Datta et al., 2005; Dharani and Aroquiaraj, 2013; Rajam and Valli, 2013; Singhai and Shandilya, 2010; Velmurugan, 2014; Yadav et al., 2014). In CBIR, query-by-example-image technique is used. In this technique,users have to provide the actual image or sketch of the actual image to the search engine to retrieve similar images. However, similarities that can be easily inferred by humans could be missed, since images and sketches are represented with low-level features like colors, shapes and textures. If image is represented with these low level features, the high level semantics are missing due to *semantic gap* (Smeulders et al., 2000).

In Text-based Image Retrieval(TBIR), images are retrieved using textual query in the same way as search text documents from the web which is an alternative technique to querying with exemplar images. In TBIR technique, texts which are associated to a required image are used to search an image from image databases on the web. These texts(or tags) are supplied by professional annotators or users manually, or semi-automatically or automatically. Manual annotation is difficult for large image databases.

Some techniques combined the two techniques(text-based image retrieval and content-based image retrieval) into one to create a hybrid image retrieval technique (Barrios et al., 2009).

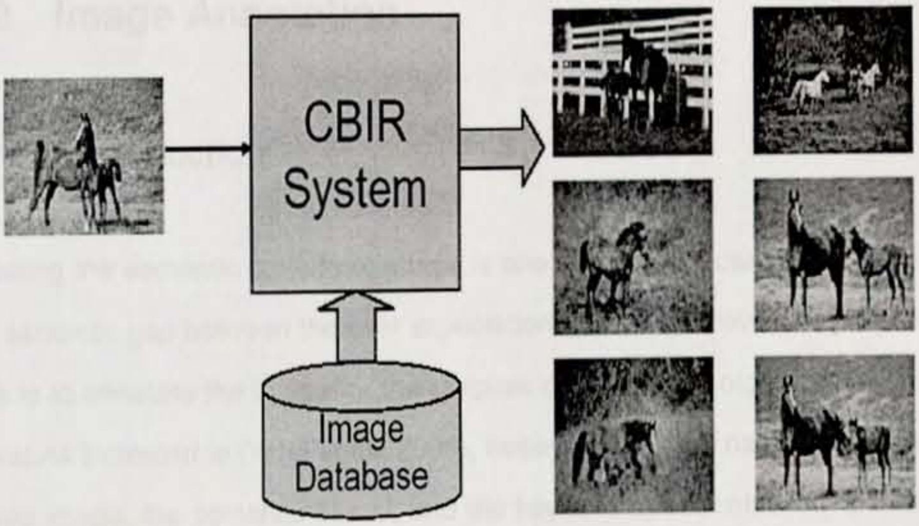

Fig. 2.2 General Content-based Image Retrieval.

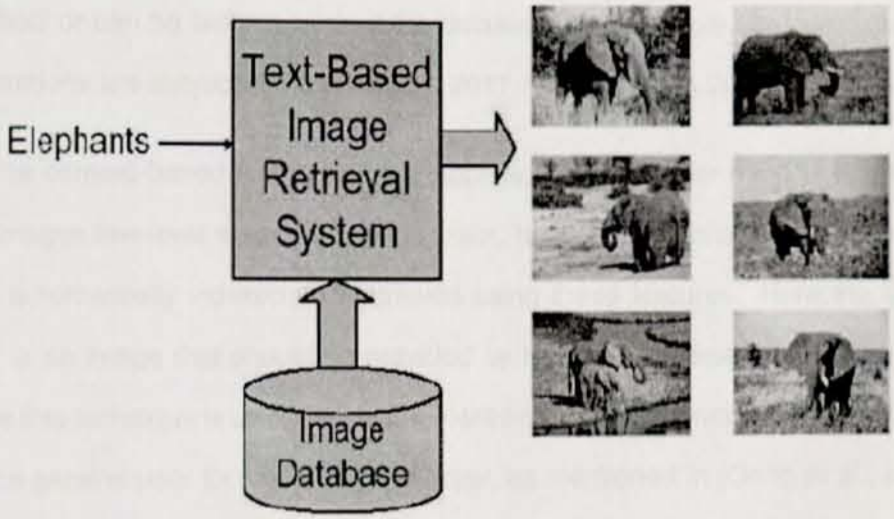

Fig. 2.3 General Text-based Image Retrieval.

2.2 Image Annotation

2.2.1 Introduction

Labeling the semantic content of image is one of the significant ideas to bridge the semantic gap between the user expectation and the low level image features. This is to annotate the image for the purpose of simplifying organization and retrieval.As indicated in (Yang et al., 2009), three approaches, namely the content-based image, the semantic-based, and the keyword-based retrievals are widely adopted to retrieve images.

In the keyword based image retrieval approach, images are annotated manually so that retrieval can be applied using similar techniques as that of text documents (Chang and Hsu, 1992). However, annotating images manually will not be practical or can be tedious work, if the dataset is large in size. Besides, human annotations are *subjective* (Jesus et al., 2011; Wenyin et al., 2001).

The content-based image retrieval approach on the other hand is based on the images low-level features such as color, textures, and shapes. Images are then automatically indexed and retrieved using these features. Here the query itself is an image that should be provided by the user (Laaksonen et al., 2000). Since this technique is using images content as a query, it is not often appropriate for the general user for retrievals. Moreover, as mentioned in (Gong et al., 2010) content-based image retrieval systems also suffer from the problem of semantic gaps - a gap between perceptual issues and conceptual issues (Smeulders et al., 2000; Wang et al., 2012).

The third approach of image retrieval is using the technique of automatic image annotation. In this technique, we can retrieve images like text documents using textual queries. In automatic image annotation technique, the model will learn automatically semantics concepts to assign meta-data to images in the form of captions and once images are annotated with semantic labels, they can be retrieved by keyword queries in similar way to that of textual document retrievals (Fan et al., 2010; Liu et al., 2009).

The main characteristic of automatic image annotation is enabling users text-based search based on image content and benefits from the use of both the text based annotation and content-based image retrieval(Zhang et al., 2012). To apply keyword-based image retrieval in large repositories effectively, automatic image annotation is necessary (Yang et al., 2009).

To address the problem of automatic image annotations, researchers introduced different techniques from which some are aimed to create a probabilistic model for the purpose of representing the correlation between images and annotation words. Here the word probability is conditioned on the image regions or image features, or the joint probability of the word and the image are computed to select an annotation word for untagged image based on tags of previously annotated images. These approaches have attracted much attention to model the image-word correlation (Gong et al., 2010) ; therefore, fewer researches are done to utilize the word-word correlation for image annotation and more should be done to exploit the tags (or words) relatedness in image annotation.

As indicated in (Gemmell et al., 2008c) , because of the development of Web2.0 (evolution from Web 1.0 to Web 2.0), there is a transition of generating contents from owners websites to a more open and social Web where users are not only the information consumers but also producers.

This new age of the Web brought diversity of new social application like Flickr Photo sharing website. Given the increase in popularity of such photo sharing websites, there is a recent research focus on the indexing and retrieval of such content (McParlane et al., 2014).

2.2.2 Image Annotation Techniques

Image annotation is mainly used to organize and manage images database for the purpose of future retrieval and sharing, in such a way that annotated images can be searched using keyword-based retrieval and can easily be communicated to other users. Whereas, it is tremendously difficult to retrieve non-annotated images from large images databases (Wenyin et al., 2001) and they may not carry some information for others in textual form. As mentioned above annotation can be done in three ways: Manually, automatically and Semi-automatically(human-machine interactive or assistive tagging). In this section we present previous works on these image annotation techniques .

Manual Image Tagging

In manual image annotations, user assigns tags to new individual or batch of photos which is a completely human oriented (Raheja and Gupta, 2011). For instance, image sharing sites like Flickr, provide for their users the facilities of manual images annotation while uploading their images for the purpose of organization, future retrieval and communication (Barai and Cardenas, 2010).

In manual annotation, carefully chosen tags for a given image can improve the accuracy of image retrieval systems. But the challenging issues of manual annotation is that it is tedious, labor intensive and subjective (Jesus et al., 2011; Wenyin et al., 2001). The promising technique to alleviate these problems is using semi-automatic or fully automatic image annotation or using a hybrid of these techniques. When we use a hybrid of the semi-automatic and fully automatic image annotation, we can incorporate the advantages of manual annotation (Zhou et al., 2011). These techniques are reviewed below in detail.

Automatic Image Tagging

The purpose of Automatic Image Annotation is to assign images' textual descriptions in the form of captions without or with limited human intervention (Bernardi et al., 2016; Zhang et al., 2012). They can be model-based and/or search-based (Wang et al., 2012).

Model-based automatic image annotation is, an approach,based on the technique of determining a one-to-one correspondence between the low-level image features to some high-level semantic concepts by analyzing previously labeled images for training set on the objective of building a model which can predict labels of the new image. Whereas in search-based automatic image annotation, to label a new image, similar images are searched and the tags of these images will be analyzed and propagated to the new image as annotation tags. Reviews of research works (on both model based and search based) techniques in the field of automatic image annotation are presented below.

The work in (Mori et al., 1999) tried to solve image annotation problem by developing *a co-occurrence model* which is useful to represent the relationship between keywords and visual features. In this annotation each image was divided into non-overlapping *tiles* by vector quantization and centroids of clusters are obtained, then words are assigned for each centroid by estimating probability for each word in each centroid statistically to that part of image. The annotation technique was the first attempts in automatic image annotation and initiated the research of automatic image annotation. In this model, for a word w and each cluster c, a score is calculated as follows:

$$P(w/c) = \frac{m_{w,c}}{\sum_w m_{w,c}}, \tag{2.1}$$

where $m_{w,c}$ is the number of times the word w co-occurs with a tile from cluster c, and

$$s(w,x) = \frac{1}{|x|} \sum_{t \in x} P(w/c_t), \tag{2.2}$$

where $s(w,x)$ is the relevance score of w for an unseen image x, c_t is the nearest cluster of $x's$ tile t and $|x|$ is the number of tiles in x.

A machine translation model, to address the problem of image annotation which improved *the co-occurrence method* by the work in (Mori et al., 1999) proposed by (Duygulu et al., 2002). The model translates image segments (or blobs) to annotating words using Expectation Maximization (EM) algorithm. This model assumed keywords and visual features as different languages. In the work the process of attaching labels to image regions is analogous to the translation of one of representations say image regions (French) to another form say labels (English) using parallel corpus, a translation from vocabulary of "blobs" to vocabulary of "labels" or "texts" using linguistic translation machine. The work tried to argue region based annotation is more interesting than global based ones, since region based annotation provides information on which part of the image is related to which label. This work was improved by (Virga and Duygulu, 2005) by comparing multiple statistical machine translation models as well as language modeling techniques to capture both the visual appearance of images and the semantic relations between annotations.

The likelihood function of the work of (Duygulu et al., 2002) is defined as:

$$P(w/b) = \prod_{n=1}^{N} \prod_{j=1}^{M_n} \sum_{i=1}^{L_n} P(a_{nj} = i) t(w = w_{nj}/b = b_{ni}), \tag{2.3}$$

where N is the number of images and M_n and L_n are the number of words and blobs associated with image n, respectively. To translate between the set of blobs comprising an image and annotation estimates translation probabilities $p(a_{nj} = i)$ in image n, a particular blob b_i is associated with a specific word w_j, that maximize the likelihood on the training set.

Blei and Jordan (2003) developed a model which can find conditional relationships between latent variable representations of sets of image regions and sets of words. They used direct co-occurrences between image region and word (word-to-image relations) in the process of annotation. They extended the Latent Dirichlet Allocation (LDA) Model and proposed a Correlation LDA model which relates words and images.

A model called, *coherent language model*, proposed by the work in (Jin et al., 2004) for automatic image annotation considered word-to-word relatedness (or correlation) as important feature in the prediction of words as annotation of a given image. In their model, they demonstrated that prediction of annotation words is no longer independent from each other. Let us assume two words w_1 and w_2. The estimation of the probability of w_1 depends on the estimation of the probability of w_2, and vice versa. The use of word-to-word pairwise correlation information is taken into account through the Expectation Maximization algorithm for finding an optimal language model for a given image.

The work in (Jeon et al., 2003) improved the work of (Duygulu et al., 2002) by assuming that image annotation as *cross-language information retrieval* problem and developed *a Cross-Media Relevance Model* (CMRM). This model treated words and clustered blobs as different lexicons. This and other relevance models like in (Feng et al., 2004; Lavrenko et al., 2003; Yavlinsky et al., 2005) tried to model the joint distribution of vocabulary of blobs and words. That means mainly they have focused on finding word-to-image relationships. They assume that a set of blobs is related to a set of words instead of assuming one-to-one correspondence between the blobs and words. They proposed a model, *Continuous-space Relevance Model* (CRM) which generalized the previous CMRM to model high dimensional continuous features without clustering and quantization.CRM improved by the work in (Feng et al., 2004). They developed a model called *Multiple Bernoulli Relevance Model* (MBRM) and carried out the improvement by computing features from segmented image regions. The model is used to compute the joint probability distribution between words and images using a multiple Bernoulli model.

Metzler and Manmatha (2004) proposed an approach, called Interface Network, able to link regions and their annotations. A new image can be annotated using this approach by propagating belief through the network to the nodes of the network which used to represent keywords.

The authors in (Jeon and Manmatha, 2004) proposed a statistical technique used to predict the probability of a label given test data using Maximum Entropy approach, for the first time in image annotation, in the process of automatic image annotation. The image is represented using a language of visterms (visual terms) which are clusters of rectangular regions and similar to terms in languages.

A technique, which used content based image retrieval is proposed by (Li et al., 2006) for automatic image annotation in such a way that, visually similar images with rich annotations (titles and surrounding texts) are searched from the web and these annotations are analyzed and ranked to annotate the new image. The authors in (Hardoon et al., 2006) used similar approaches but applied Scale-Invariant Feature Transform (SIFT) algorithm-an algorithm to detect and describe local features in images (Lowe, 2013), as features extractor.

A technique proposed by the work in (Liu et al., 2007a) to solve the problem of image annotation involved two types of relations: one is the word-to-image relation and the other is the word-to-word relation. Both relations can be estimated by using search techniques on the web data as well as available training data. Besides estimating word-to-image relations the model estimates the joint probability by the expectation over words in a pre-defined lexicon to find word-to-word relations.

An automatic image annotation technique using language model, called the semantic similarity language model, to improve the performance of the existing image annotations in (Duygulu et al., 2002; Jeon et al., 2003) by utilizing probabilistic models is proposed by (Gong et al., 2010). The model used to estimate the semantic similarity among the annotation words. The authors claimed and showed annotations tags that are more semantically coherent will have higher probability to be chosen in the annotation process.

Some works used partially annotated images (users supplied tags) to automatically expand tags. (Sevil et al., 2010) proposed tags expansion systems which can automatically expand tags of images using both textual and visual features of previously annotated and related images and the work in (Sigurbjörnsson and Van Zwol, 2008) used tag co-occurrence statistics in order to recommend annotations for partially tagged photos. Similarly, (Lindstaedt et al., 2009) proposed image annotation techniques to propagate user generated folksonomic annotations to deal with unlimited vocabulary.

The authors in (Lin et al., 2012b) proposed image auto-annotation model, named TagSearcher, to predict tag scores by considering weights of visual neighbors, votes for candidate tags and tag specific trust-estimated with respect to a candidate tag using graphical model(tag-related random search optimization algorithm). TagSearcher proposed to use a constrained range of visual neighbors for label propagation, and utilizes tag-related random search processes to find out the trustworthy part for each candidate tag.

The annotation process implemented in (Jeon et al., 2003) used a set of annotated images and could learn the joint distribution of the blobs and concepts. The blobs are clusters of image regions obtained using the K-means algorithm. Using discrete sequence of blobs identifiers the set of blobs of each image from the test set represented and the distribution used to generate a set of concepts for a new image.

Blei and Jordan (2003) presented Multi-Modal Hierarchical Aspect Model and Mixture of Multi-Modal Latent Dirichlet Allocation model for automatic image annotation based on Hofmann's hierarchical model for text and the extension of LDA, respectively.

The authors in (Chen et al., 2010) argued about on the relation of each word topic and multiple visual topics and said each word topic related to multiple visual topics, with different connection strength respectively. Based on this assumption, the authors proposed a probabilistic Topic-Connection (PTC) model for automatic image annotation. The model was estimated via collapsed Gibbs sampling algorithm, while the parameter selection was done by studying the likelihood and perplexity. This model was compared in terms of performance with the Corr-LDA model under the same automatic image annotation scenario using cross-validation and better results are obtained.

A technique, which derived semantic correlation matrix from Flickr's related tags resource to model conditional random field for web image annotation is proposed by (Xu et al., 2009). The model integrates keyword correlation derived from Flickr, and the textual and visual features of images to boost the performance of the annotation.

A framework of using language models, called the semantic similarity language model, to improve the performance of Machine Translation model and Cross Media Relevance Model,is proposed by (Gong et al., 2010). They added language model to the original image annotation models, and the overall annotation performance improved as the annotation word bias reduced and the system is more likely to generate semantically coherent annotation word set.

Horiuchi et al. (2013) presented an automatic image description which used word level features. In their techniques, similar images of the query image are searched, and phrase level frequencies and similarity ratios as the word level features are computed. Finally, general nouns and their attributes(used to form the phrase) based on word level features are extracted and ranked to have descriptions of the target image. They presented the general framework for the description as shown in Figure 2.4 by (Horiuchi et al., 2013)

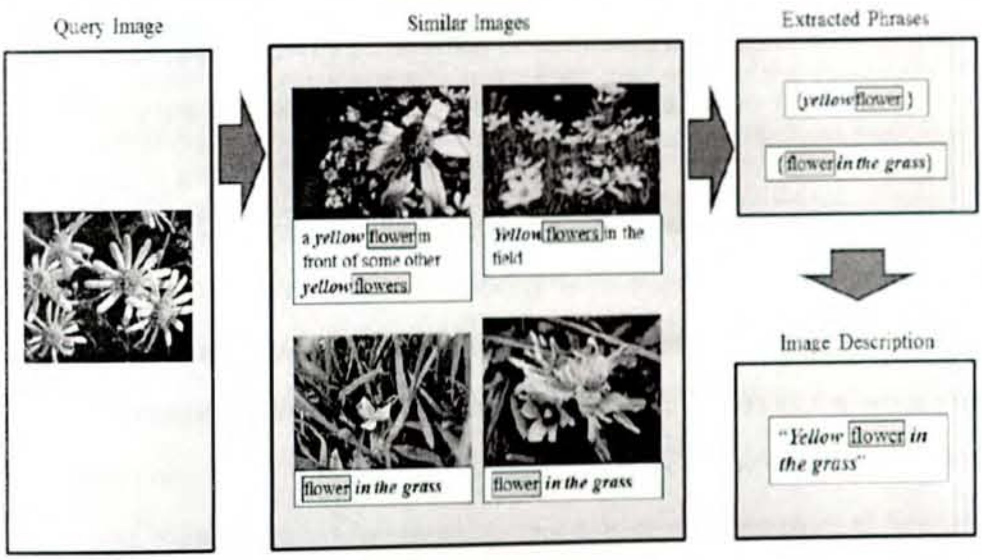

Fig. 2.4 Pipeline of image description

The paper in (Tiwari and Kamde, 2015) proposed a system for automatic image annotation along with image retrieval. It is designed basically for tagging web images and retrieving. In the process of tagging, the system extracts contextual information from web pages and important keywords from the document are extracted using text processing.Terms obtained are assigned as tags based on their corresponding weights. The authors in this work tried to bridge the semantic gap by exploiting web contextual information related to the image.

Ghosh and Bandyopadhyay (2013) and Leong et al. (2010) used text mining technique for automatic image annotation to assign relevant tags to the new(or unseen) image. These techniques are implemented using the classification logic of the text mining algorithm that assigns the given set of keywords or tags to some predefined classes.

An approach proposed by the work in (Kamoi et al., 2007) improved the accuracy of image recognition by considering word-to-word correlation between words representing the context and the objects of the image and two semantic levels of keywords that give feedback to each other. The work in (Llorente et al., 2008) hypothesized more accurate annotations can be produced by introducing additional knowledge in the form of statistical co-occurrence of terms which is provided by the context of images. According to the work in (Gong et al., 2010) much attention is given to model the image-word correlation and tried to address word-to-word direct relations.

The paper presented in (Ke et al., 2012) proposed a hierarchical image automatic annotation model by establishing links between images and key words (word-to-image relations) and combining the advantages of discriminative models and generative models. Here word-to-word relations are not well explored.

The authors in (Feng and Lapata, 2008) tried to show the possibility of developing an annotation model using images embedded in news articles. They tried to leverage the vast resource of images available on the Internet while exploiting the fact that many of them are labeled with captions.

The paper presented in (Barai and Cardenas, 2010) is more related to our study. This paper presented an automatic image annotation techniques called I-Tag and used both visual and textual information of the images and recommends relevant tags to an image. It used state of the art tools on text based retrieval and content based retrieval to retrieve similar images to the target image. Lastly, frequently occurring tags obtained from previously annotated images are recommended as tags of the new image.

Semi-automatic Image Tagging

In a semi-automatic image annotation, the accuracy of manual annotation and the efficiency of automatic image annotation are combined (Raheja and Gupta, 2011). Here under, previous works which assume that the target image is annotated partially in manual and extending the initial annotations with different techniques are presented. In our work, initial tag(s) which are provided by the users, when they are uploading their images for the purpose of organization and later retrieval , would be extended for the purpose of tags enhancement(or enrichment).

A model for tag(s) recommendation, by exploiting tags co-occurrences is developed by the authors in (Wu et al., 2009). In their model, an image with set of initial tag(s) would have a set of recommended tags which may have a semantical or visual correlation to the images. They used an amalgamation of three kinds of correlation analyses(measures) to rank the tags: tag co-occurrence by adopting the work in (Sigurbjörnsson and Van Zwol, 2008), tag visual correlation using the visual language model (VLM) algorithm((Wu et al., 2007)), and image conditioned tag correlation to find the similarity of tags with respect to the target image based on tags co-occurrences by using VLM. The three kinds of correlations are amalgamated using an algorithm introduced by the work in (Freund et al., 2003).

The work in (Thangam and Angel, 2013) proposed a semi-automatic photo annotation scheme which could modulate the manual efforts and the tagging performance in a flexible way. In their proposed work, features are extracted and the patters are framed to compute similarities from collections of photos. Based on these similarities exemplars are selected and users will provide tags for these exemplars manually. Based on the tags of these exemplars, the rest of photos are annotated automatically.

Social Image Tagging

With the launch of social media sites, *social (or collaborative) tagging* is becoming an easy and promising solution for ordinary users to annotate, organize and share digital content on the web (Font et al., 2013). *Social tagging* is the task of linking user-defined keywords to a multimedia object (Lin et al., 2012a; Spyrou and Mylonas, 2016).

Social resources organization and sharing systems allow ordinary users to upload, organize and share their resources. These systems(or websites) enable users to label their resources using freely chosen texts (or words) as labels or tags and users describe their contents(for instance image contents) using set of these tags (Qian et al., 2014).

There are different types of sharing systems for different purposes. For instance, Flickr[1] for organizing and sharing of images (or Photos), YouTube[2] for the sharing of videos, del.icio.us[3] for the sharing of bookmarks, CiteULike[4] for the sharing of bibliographic references, and Last.fm[5] for the sharing of music listening habits.

The working strategies of these systems are similar, that is users can have an account and using this they can upload resources to organize and share (can upload resources and tag them for the purpose of organization, future retrieval and communication). Given a user defined tags for the new image she/he is uploading, some additional image which will increase the semantics of the original tags can be predicted using statistical techniques to leverage the tag-to-tag correlation from semantically rich textual corpus (Zhou et al., 2008) and this process is commonly known as collaborative or social image tagging (Zhou et al., 2011).

In the aforementioned social tagging systems, the collection of a given user's tags assignment is termed as personomy, and collection of all personomies for a given social tagging system is called foloksonomy (Hotho et al., 2006a; Jäschke et al., 2007), which is an easy to use, developing and iterative system (Gupta

[1] https://www.flickr.com/
[2] https://www.youtube.com/
[3] https://del.icio.us/
[4] http://www.citepulike.org/
[5] https://www.last.fm/

et al., 2010). Reviews of some works about image annotation in social, collaborative, or Folksonomy Systems are presented below.

Luo and Zincir-Heywood (2004) presented two tag recommendation algorithms: an adaptation of user-based collaborative filtering and a graph-based recommender built on top of FolkRank-a ranking algorithm for folksonomies (Hotho et al., 2006a), an adaptation of the well-known PageRank algorithm (Pasquinelli, 2009) that can cope with undirected triadic hyper-edges.

There are three elements in folksonomies: tags, resources and users (Wetzker et al., 2010) and an algorithm called FolkRank based on the PageRank algorithm for ranking is provided by (Hotho et al., 2006a). FolkRank gives higher rank to the resources which are tagged by important users and important tags. The tests were performed on a dense core of folksonomy, thus might be not representative of the whole folksonomy in which some sets can be scarce in resources especially tags set.

The personomy reflects the user's vocabulary, preferences, interests and knowledge. The work in (Jäschke et al., 2007) suggested and evaluated potential sources of recommended tags, focusing on folksonomies oriented towards individual users. These suggestions are used to propose a three-step tag recommendation system. Basic tags are extracted from the resource title. In the next step, the set of potential recommendations is extended by related tags proposed by a lexicon based on co-occurrences of tags within resource's posts. Finally, tags are filtered by the user's personomy.

A hybrid approach to automatic image labeling which combines the low-level content and the high-level semantic tags provided by users to exploit advantages of both content-based and collaborative techniques is presented by (Zhou et al., 2008). The collaborative technique is used to solve the tag sparsity problem in favor of better estimation of the tag co-occurrences probabilities.

Lindstaedt et al. (2009) applied two techniques based on image analysis. The first is using controlled vocabulary to classify images and the second is for the tag propagation using user generated folksonomic annotations. The paper used the advantages of both visual content and existing folksonomies for automatic image annotation techniques.

Using GPS coordinate to locate landmark images is a common technique to tag image with the tags of similar images in the corpus. The work in (Abbasi et al., 2009) proposed a method of identifying landmark images without having GPS coordinates on the image but using tags and social Flickr groups. The method contains two main parts: exploiting tags and social Flickr groups to train a classifier to identify landmark photos and tags and ranking all suggested relevant tags by their representativeness of a landmark.

Integrating both low level and high level image features is one of the ideas of bridging the gap between the user expectation and the low level image features. As shown in (Zhou et al., 2011) a hybrid probabilistic model is used to integrate low-level image features and high-level user provided tags (or folksonomies) to automatically tag images. For images without any initial tag, the model predicts new tags based on the low-level image features. For images with user provided tags, the suggested model jointly exploits both the image features and the tags in a unified probability framework to recommend additional tags to label the images.

The work presented in (Abbasi and Staab, 2008) suggested a Triple Play method to improve search in collaborative tagging systems to generate Folksonomies like Flickr or del.icio.us particularly when there are more tags in the search query. They created a vector space model (SmoothVSM Dense or SmoothVSM Sparse) considering user-tag relationship information available in collaborative tagging systems and then apply Latent Semantic Analysis for retrieval of resources from these vector space models.

Mathes (2004a) on the other hand a text categorization system based on the combination hierarchical SOMs encoding architecture and the designed kNN classifier able to find sequences of word/word co-occurrences as well as their frequencies.

The paper in (Gemmell et al., 2008a) proposed a method for personalizing search and navigation in folksonomies based on three clustering techniques: Hierarchical Agglomerative Clustering, Maximal Complete Link Clustering and k-means Clustering. Tag clusters are used to bridge the gap between users and resources, offering a means to infer the user's interest in the resource. In this paper the authors introduce a methodology for automatically ranking and classifying photos according to their attractiveness for folksonomy members.

2.3 Dis(similarity) Metrics

In the literature, the Jensen-Shannon Divergence (JSD) (Manning and Schütze, 1999) is a widely used metrics and has shown to outperform other metrics (Ljubešić et al., 2008) under many respects, when computing the dissimilarity of probability distributions. It is based on the Kullback-Leibler Divergence (KLD) (Kullback and Leibler, 1951), however, it is symmetric and has always a finite value. Since in practice tag probability distributions are created from samples, and are necessarily affected by statistical fluctuations, in (Mousselly-Sergieh et al., 2014) a metrics based on a Maximum Likelihood estimate of the JSD was proposed, the Adapted Jensen-Shannon Divergence (AJSD), which takes into account fluctuations and provides a measure of the statistical error of the results. Before introducing the new metrics, we review the KLD and JSD approaches to calculate the distance between probability distributions.

Let us consider two tags $t_1, t_2 \in T$ and the corresponding empirical co-occurrence probability distributions $P(\mathscr{F}|t_1)$ and $P(\mathscr{F}|t_2)$ over the feature set $\mathscr{F} = \{f_1, \ldots, f_m\}$. We can simplify the notation as follows: $P(\mathscr{F}) \equiv P(\mathscr{F}|t_1)$ and $Q(\mathscr{F}) \equiv P(\mathscr{F}|t_2)$; the values of P and Q at a specific feature $f \in \mathscr{F}$, will hereafter be represented simply by $P(f)$ and $Q(f)$, respectively.

Kullback-Leibler divergence

The most typical metrics for dissimilarity between two probability distributions is the Kullback-Leibler divergence D_{KL}, defined as follows:

$$D_{KL}(P||Q) = \sum_{f \in \mathcal{F}} P(f) \ln \frac{P(f)}{Q(f)} \tag{2.4}$$

Notice that the expression $D_{KL}(P||Q)$ is asymmetric in its arguments, i.e in general $D_{KL}(P||Q) \neq D_{KL}(Q||P)$. This problem can be solved by adopting, as a definition of divergence, a symmetrized version of the previous expression:

$$D_{SKL}(P||Q) = \frac{1}{2} \left\{ \sum_{f \in \mathcal{F}} P(f) \ln \frac{P(f)}{Q(f)} + \sum_{f \in \mathcal{F}} Q(f) \ln \frac{Q(f)}{P(f)} \right\} \tag{2.5}$$

However the SKL divergences become infinite as soon as either P or Q vanish in one point of the support set.

Jensen-Shannon divergence

This problem can be fixed by using the Jensen-Shannon (JS) Divergence, which is given by the following equation

$$\begin{aligned} D_{JS}(P||Q) = \frac{1}{2} \sum_{f \in \mathcal{F}} \Bigg(& P(f) \ln \frac{2P(f)}{P(f)+Q(f)} \\ & + Q(f) \ln \frac{2Q(f)}{P(f)+Q(f)} \Bigg) \end{aligned} \tag{2.6}$$

which differs from the SKL divergence of equation (2.5) in that the denominator of the logarithm's argument consists now in the arithmetic average $(P(f)+Q(f))/2$ of the two functions.

Adapted Jensen-Shannon Divergence

If, as in our case, the probabilities P and Q are not available, we have an estimate of them through a finite sample represented in the form of a histogram for P and a histogram for Q. In this case the divergence computed on the histograms is a random variable. This variable, under appropriate assumptions, can be used to compute an estimate of the divergence between P and Q using error propagation under a Maximum Likelihood (ML) approach, as illustrated in (Moussselly-Sergieh et al., 2014). Hereafter we give the practical definition for the computation of the two divergences JSD and AJSD.

Given two histograms, their JSD and AJSD divergences can be computed as follows.

Given the arrays

$$x_f = P(f) = \frac{k_f}{n} \qquad y_f = Q(f) = \frac{h_f}{m}$$

representing two probability distributions over the same set of indexes $f \in F$, the Symmetrized Kullback-Leibler Divergence is defined by

$$d_{SKL}(x||y) \equiv \frac{1}{2}\sum_f \left(x_f \ln \frac{x_f}{y_f} + y_f \ln \frac{y_f}{x_f} \right)$$

Given the arrays x_f and y_f as above, the Jensen-Shannon divergence is defined as

$$d_{JS}(x,y) \equiv \frac{1}{2}\sum_f \left(x_f \ln \frac{2x_f}{x_f + y_f} + y_f \ln \frac{2y_f}{x_f + y_f} \right)$$

Let us define z_f as follows

$$z_f \equiv \left(x_f \ln \frac{2x_f}{x_f + y_f} + y_f \ln \frac{2y_f}{x_f + y_f} \right)$$

We take x_f and y_f as estimates of true quantities and we approximate their variances by

$$\sigma^2(x_f) = \frac{k_f}{n^2} \qquad \sigma^2(y_f) = \frac{h_f}{m^2}$$

then

$$\sigma^2(z_f) = \left(\ln \frac{2x_f}{x_f + y_f}\right)^2 \sigma^2(x_f) + \left(\ln \frac{2y_f}{x_f + y_f}\right)^2 \sigma^2(y_f)$$

now compute the weights $w_f = \left(\sigma^2(z_f)\right)^{-1}$. Finally the value of the Adapted Jensen-Shannon divergence is computed as

$$w(z_f) \propto \frac{\sum_f w_f z_f}{\sum_f w_f}$$

2.4 Summary of the Chapter

The objective of annotating images is to make them searchable using textual information as text documents. Assigning tags to images helps to improve images retrieval on the web (Trant, 2009) because content-based image retrieval suffers with the problem of the semantic gap. However, we can allow a high-level, concept-based retrieval by bridging the semantic gab between the low-level image representation and the high-level interpretation. This can be achieved by assigning meaningful tags (words) to images through the methods of manual, Semi-automatic and automatic annotation techniques.

Acquiring these tags/words to images using manually labeling technique is difficult since manual labeling is time consuming, tedious, and subjective. The solution to this is generating images' tags based on the visual content automatically. Even if automatic image annotation is fast to assign tags to images it suffers with the known semantic gap problem. Current trends show that alternative solution is leveraging both as a hybrid annotation technique and introducing social, or collaborative or folksonomy systems. In the literature above we have tried to present the state of the art of Image Annotations in these areas of interest. Lastly, common probability distribution distance measures used in the literature to measure the similarity or dissimilarity between two divergences are presented with their importance one on the other.

Chapter 3 Related Works

In this chapter, research efforts closely related to this study, including tag sense disambiguations, tags relatedness analysis with the corresponding metrics used and tags' features selection techniques are explored.

3.1 Tags Disambiguation in Collaborative tagging

Imprecise or ambiguous tag assignments, due to the freedom of using freely chosen texts or tags and uncontrolled manner of tags creation in Folksonomy, can decrease the performance of tagging systems regarding search and retrieval of relevant items(Abel et al., 2010).Tag sense disambiguation can be applied to the vocabulary of social tags, thereby enabling future use and retrieval simple.

Tag/word sense disambiguation is to attribute the correct senses or provide enough information to tags/words in a given context. It is regarded as one of the most interesting and longest-standing problems in information retrieval since, ambiguous tags in the queries are problematic for information retrieval systems.

Tags/words disambiguation can be done automatically or based on human judgement. In automatic tag disambiguation, for a user provided initial tag, the system will assign its meaning (set of possible senses) without human intervention. In user assisted tag disambiguation process, the system will provide set of possible senses of an initial user provided tag and the user will select one or more in his/her mind for a specific context. These senses can be computed from a given corpus automatically based on tags relatedness analysis.

As indicated in (Giyanani, 2013; Navigli, 2009; Pal and Saha, 2015) main approaches to word sense disambiguation can be either knowledge-based(based on different knowledge sources like machine readable dictionaries or sense inventories, thesauri, etc) or corpus-based. In addition, the latter can further be categorized as supervised ones and unsupervised ones:

1. **Supervised tag/word sense disambiguation:** in this technique, machine learning strategy is used to learn a tags classifier algorithm from carefully sense-annotated corpora(Martín Wanton and Berlanga Llavori, 2012).

   This approach suffers from the lack of readily available manually semantically sense-annotated corpora, from which to construct really inclusive systems. Moreover, in social tagging systems, users are free to choose any tag for their resources and for such freely chosen tags, readily sense-annotated corpora are difficult to get. This is the basic reason many unsupervised approaches being studied by many researchers in word sense disambiguation (Chaplot et al., 2015).

2. **Unsupervised tag/word sense disambiguation:** this technique does not depend on sense-annotated corpora i.e. it does not rely on external knowledge source such as machine readable dictionaries, concept hierarchies, or sense tagged text. The technique do not use any manually sense-annotated corpus to suggest senses of a tag for a specific context. This technique assumes that tags relatednesses (i.e. similarities and dissimilarities of tags)

are important in determining tag senses in a given context using un-annotated corpus (Mante et al., 2014; Martín Wanton and Berlanga Llavori, 2012).

Mostly, tag sense disambiguation techniques in folksonomies belong to the unsupervised tag sense disambiguation approaches since selection of senses is free i.e. it does not depend on external knowledge, representation of context is not depend on a given context window or paragraph for a target tag and knowledge sources is unavailable, unlike word sense disambiguation.

Since we want to disambiguate a user provided initial tag in a typical semi-automatic image tagging our approach belongs to user assisted unsupervised tag sense disambiguation technique which is slightly different from automatic word sense disambiguation. In our case user will be involved to determine the appropriate sense for the initial tag from suggested lists of cue-words.

In order to determine cue-words in tag-sense disambiguation process, researches tried different techniques for determining related tags by exploiting tags usage patterns. Commonly, they focused on identifying groups of tags that have similar semantics by tag co-occurrence statistics and clustering techniques. Tags relatedness information is used to determine the possible meaning of individual tags. Techniques used to determine tags/words relatedness information in finding set of possible senses of a given tag are reviewed in section 3.2. These techniques can be also used in determining tags co-occurrence for automatic image's tags recommendation procedure(section 4.4).

3.2 Tags Relatedness Using Co-occurrence Analysis

First a clarification is in order, on the use of the terms *similarity* and *relatedness*. Semantic *similarity* and semantic *relatedness* are two linked concepts, but are not synonyms. (Budanitsky and Hirst, 2006) point out that relatedness is a more general concept than similarity: similar entities are semantically related via their similarity ("auto"-"car"), but non-similar entities may also be semantically related by meronymy, a.k.a. part-of relationship ("hand" and "palm"), antinomy ("left" and "right"), rather than just frequent association. Applications typically require relatedness rather than similarity: for example, "leaf" and "hand" are cues which can be used to disambiguate the term "palm". In the work, the term *dissimilarity* is used as an antonym of the term *relatedness*.

Hereafter, which research works tried to exploit tags/words semantic relatedness in determining tag/word possible senses in tag sense disambiguation process are presented. Moreover, research works techniques used to determine co-occurrence of tags to improve semantics are also reviewed.

The work presented in (Purandare and Pedersen, 2004), proposed a clustering approach which using the agglomerative clustering technique in determining words relatedness for the word-sense disambiguation. In this technique, for each word co-occurrence a single cluster is constructed. The agglomerative clustering technique is used to merge the most similar pair of clusters and continued to the less similar pairs until a halt threshold obtained. Given the target word(tag), the method which is based on word clustering to identify words that are more similar to the target word and use these clusters to have specific senses has also been proposed in the study (Pantel and Lin, 2002). The proposed approach, called the clustering by committee which applied a different word clustering method, is an extension of the work presented in (Lin, 1998). Purandare and Pedersen (2004) selects features based on their frequency counts or log-likelihood ratios in corpus used.

The work in (Cattuto et al., 2008) presented different types of tags relatedness measures such as co-occurrence count, distributional measures based on cosine similarity and graph-based measure adapted from PageRank (Brin and Page, 2012) called FolkRank (Hotho et al., 2006a) to folksonomies.The goal is to assess how a given tag relatedness measure approximates a reference measure. By taking snapshot of social bookmarking system, del.icio.us, the work presented in this (Cattuto et al., 2008), mapped the tags of this system to synsets of WordNet and use the semantic relations of WordNet to infer corresponding semantic relations in the folksonomy. To build most important features set frequency counts of most frequently co-occurring tags is used.

Llorente and Rüger (2008) showed more accurate annotations can be generated by presenting additional knowledge in the form of statistical co-occurrence of words. This approach is useful to correlate the annotation words with the context of the images and the work in (Llorente et al., 2009) words co-occurrence statistics at image level is used to annotate images. A similar approach is used to increase the accuracy of annotation words for unseen image by the work in (Liu et al., 2007a). Besides estimating word-to-image relations the model estimates the joint probability by the expectation over words in a per-defined lexicon to find word-to-word relations.

The work in (Lee et al., 2009) proposed a method, named Tag Sense Disambiguation, which used to map a target tag to a topic in a Wikipedia articles by finding its local and global neighbor tags based on co-occurrence relations. The resulting tags are used as contexts of the target tag based on the topic relevance values between this context and all the Wikipedia topics measured based on the term frequency of Neighbor tags on a Wikipedia text.

The work in (Begelman et al., 2006) a cut-off threshold, determined using the first and the second derivatives of the tag co-occurrence curve, is used to decide whether two tags are related or not. The study used the frequency counts of all the co-occurring tags pairs and attempted to identify the importance of co-occurring tags just by determining the pairs of tags that co-occur significantly more frequently then expected. Then, tag clusters are built by providing the computed tag similarity matrix as input to a spectral bisection clustering algorithm.

For their tag clustering approach, (Specia and Motta, 2007) propose to represent each tag by a co-occurrence row vector and compute the similarity between two tags by applying the cosine metrics on the corresponding vectors. To determine cluster of the highly co-occurring tags, a similarity threshold to filter out pairs of tags that are not highly similar is used.

Simpson (2008) proposed a tag relatedness approach which uses the Jaccard metrics to normalize tag co-occurrences. The tags are then organized in a co-occurrence graph, which is then fed to an iterative divisive clustering algorithm to identify clusters of related tags.

The tag relatedness metrics presented by (Papadopoulos et al., 2010), is based on a graph-theoretical metrics. Tags are organized in a graph with the edges weighted according to the structural similarity between nodes: tags with a large number of common neighbors are considered related.

Tag(s) recommendation system which is based on users provided initial tag(s) proposed by the work in (Sigurbjörnsson and Van Zwol, 2008) is more related to our work. For a new image with set of user-defined tags, an ordered list of a defined number of candidate tags is extracted for each of this set of user defined tags, based on the statistical information obtained from tags co-occurrences. These co-occurrences are measured using a symmetric measure based on Jaccard coefficient and asymmetric measure to normalize the co-occurrence of a tag pair based on the number of occurrences of a tag in the tags pair. To merge these lists into a single ranking, they used tag aggregation and promotion techniques using voting and summing. This work tried to improve the semantics of tags by exploiting tags co-occurrences. The paper mainly used Jaccard coefficient to find tags relatedness. This measures is useful to find similarities using simple matching. Important features are selected based on number frequently occurring tags. In our work we have used AJSD measure which is adapted from the known JSD measure. These divergence measures used to measure the distance between two probability distributions . The AJSD measures not only average distance(such that transforms both vectors into probability distributions, builds a mean distribution and measures the mean divergence of both original distributions to the mean distribution) but also considers fluctuations of number of elements and their values(i.e.considers statistical fluctuations). Another limitation of the work presented in the aforementioned paper is the problem of ambiguity. Tags can face the problem of ambiguity which is a common problem in social tagging systems and it is not addressed but we have addressed the issue of ambiguous tags. A user may not known the exact name of a given image and he/she may submit it without tag. This issue is also not addressed in the work. in our work we have used automatic image tagging strategy to recommend tags of image which would be uploaded without tag.

Weinberger et al. (2008) proposed a statistical approach for identifying ambiguous tags based on the Kullback-Leibler (KL) divergence. For this purpose, a representation for each tag is created based on the co-occurrence with only the top frequent tags in the folksonomy.

All the above works start by exploiting tag co-occurrence counts to define their tag relatedness metrics to determine tags similarities and/or dissimilarities. They have used tags relatedness in solving the problem of tags ambiguity and to improve image annotation tags accuracy, clustering accuracy. Following using tags co-occurrence counts, either a simple threshold for tag co-occurrences like in (Begelman et al., 2006; Simpson, 2008) or the cosine metrics are used to identify similar tags like in (Cattuto et al., 2008; Gemmell et al., 2008b,d; Specia and Motta, 2007) or Jacard Coefficient to determine tags co-occurrences by the work in (Sigurbjörnsson and Van Zwol, 2008) . The Adapted Jensen-Shannon Divergence with respect to the literature, brings original contributions in the following respect: although it uses the similar representation for tags as histograms as in (Weinberger et al., 2008), it takes into account statistical fluctuations, which is crucial in any typical dataset.In the work of (Mousselly-Sergieh et al., 2014), where it was first proposed, the AJSD metrics was assessed using the WordNet corpus(relations in WordNet do not cross part of speech boundaries and some tags/words do not found in the wordNet), in terms of correlation with the Jinag & Conrath semantic similarity measure (Jiang and Conrath, 1997), which, in turn, shows a high correlation with human judgment. In this thesis, we test the metrics in the specific application context of disambiguation and use the controlled conditions provided by the synthetic corpus to perform a precise measurement of the quality of the results and use it to the real world dataset.

3.3 Feature Selection

Dimensionality reduction is a known technique to remove noisy, irrelevant and redundant features(that is to reduce curse of dimensionality). Curse of dimensionality is defined by the authors in (Li et al., 2016) as:

> *"...the phenomenon that data becomes sparser in high-dimensional space, adversely affecting algorithms designed for low-dimensional space."*

Dimensionality reduction techniques can be categorized mainly into features extraction and features selection (Ghodsi, 2006; Guyon and Elisseeff, 2003; Kim et al., 2005; Saeys et al., 2007; Van Der Maaten et al., 2009; Yan et al., 2006).

Feature extraction approaches project features into a new feature space with lower dimensionality and the new constructed features are usually combinations of original feature (Guyon and Elisseeff, 2003; Thangavel and Pethalakshmi, 2009; Van Der Maaten et al., 2009). Some of known features extraction techniques are Singular Value Decomposition(SVD) (Golub and Van Loan, 2012), Canonical Correlation Analysis(CCA) (Hardoon et al., 2004),Linear Discriminant Analysis (LDA) (Chen, 2005) and Principle Component Analysis(PCA) (Holland, 2008).

Features selection methods can be broadly categorized into two: Supervised and Unsupervised features selections (Banerjee and Pal, 2015; Qian and Zhai, 2013; Roffo, 2016). In supervised features selection technique class labels are used to determine features set for the task on hand, whereas in unsupervised features selection, there are no class labels to guide search elements of the features set (Alelyani et al., 2013; Guyon and Elisseeff, 2003; Leung et al., 2006). The new tags/words feature selection strategy presented in this study is Unsupervised Features Selection techniques, since the procedure did not consider pre-labeled classes for text classification.

The features selection approaches aim at selecting a small subset of features that minimize redundancy and maximize relevance to the target problem (Chandrashekar and Sahin, 2014; Kumari, 2012; Liu and Motoda, 2007; Saeys et al., 2007; Tang et al., 2014; Wu et al., 2010) either in backward elimination or forward selection of individual elements of the feature set, and their generalized form. Unlike feature extraction, this technique is used to select a subset of the original features set without performing feature transformation (Alelyani et al., 2013; Belanche and González, 2011; Bolón-Canedo et al., 2013; Yu and Liu, 2003). Within this perspective, it is defined by (Wang et al., 2017) as follows:

> *"Feature selection, as a dimensionality reduction technique, aims to choose a small subset of the relevant features from the original ones by removing irrelevant, redundant or noisy features."*

Examples of some of known feature selection techniques are Fisher Score (Duda et al., 2012), Spectral Feature Selection (Zhao and Liu, 2007), Information Gain (Cover and Thomas, 2012), Laplacian Score (He et al., 2006),Unsupervised Quick Reduct algorithm, Empirical Distribution Ranking, and Expectation-Maximization clustering (Dy and Brodley, 2004), to name a few.

The importance of features set selection is to have a small number of representative features which can avoid curse of dimensionality, reduce feature measurement cost, computational weight and improve accuracy(or performance) of procedures on hand (Guyon and Elisseeff, 2003; Parveen et al., 2012; Ruiz et al., 2009; Tang et al., 2014). More detailed discussions on Feature Selection is found in (Tang et al., 2014).

In the Shapley Value based feature selection used here in this thesis, an individual feature will be selected based on its degree of "cooperativeness" with other features within the procedure at hand (Serrano, 2007). The procedure proposed is basically centered in finding tags/words relatedness. The approach is used as a feature selection, for instance, in the work (Afghah et al., 2015; Cohen et al., 2005) within standard classification problems, and others used classical feature selection techniques from a statistical perspective, classical methods lose their properties (Balbi, 2010).

To the best of our knowledge, Shapley Value based features selection has never been used in tags/words classification to find their relatedness toward similarity and/or dissimilarity based on co-occurrence statistics applied in tags disambiguation and automatic tags recommendation. Almost all of the works mentioned under the above section 3.2 used frequency counts to determine best features based on the actual counts of co-occurrences in a corpus.The effectiveness of these methods depends on the coverage and characteristics of the corpus used. In our case we have used an approach which takes into account the marginal contribution of each of the tags toward the analysis of relatedness which it participates. These approach is discussed in section 3.3.1

3.3.1 Shapley Value Analysis

The Shapley Value analysis, which was introduced by Lloyd Shapley (Shapley, 1953), can detect irrelevance and redundancy by computing the average marginal contribution of a word (a channel) to the overall characterization of a context. After detecting irrelevant or redundant feature words, one can reduce the feature set to improve efficiency and quality of the classification procedure (which in turn improves the disambiguation results and as well as set of tags co-occurrence list)

In cooperative game theory there is the concept of coalitional games. In coalitional games,set of players is related to a real valued function that determines the benefit which will be achieved by sub-coalitions in the game (Cohen et al., 2005).

A coalitional game is defined by a pair (N,v) where $N = \{1,2,...,n\}$ is the set of all players and v is a real valued set function associating a worth with the coalition C, for every $C \subseteq N$.

The contribution of each tag(or word) or player to the classification task (to the disambiguation process-for instance in finding pairs of cue-words which are different in context each other and do not co-occur together in sentences but co-occur with the target word) is estimated based on the shapely value.

Definition: The marginal contribution (or importance) of player f to a coalition C, with $f \notin C$ is

$$\Delta(f,C) = v(C \cup \{f\}) - v(C) \tag{3.1}$$

The Shapley value of a player f in a game is defined by:

$$SV_{(v)}(f) = \frac{1}{n!} \sum_{\pi \in \Pi_N} \Delta(f, C_f^q(\pi)) \tag{3.2}$$

where Π_N is the set of the permutations of N and $C_f^q(\pi)$ is the coalition consisting of the set of players appearing before the f^{th} player in permutation π. In other words, the Shapley Value of a player is a suitably weighted mean of its added value to all possible subsets of players.

3.4 Summary of the Chapter

In this chapter works which are more related to this work are presented. As presented in the review , works done so far to analyze tags/words co-occurrences based on different types of metrics are discussed. Most of them used Jaccard coefficient(more about this metric is found in (Niwattanakul et al., 2013), Cosine

Similarity metrics, Kullback-Leibler divergence, and a simple Jensen-Shannon Divergence. In this thesis the Adapted Jensen-Shannon Divergence, a metric which considers the variability of values of attributes of objects in finding the distance between their probabilistic distributions of co-occurrences, is used.

Moreover, we have used a feature selection metrics in the proposed procedures-Shapley Value Calculator, to determine the marginal contribution of each tag or word toward the analysis of tags relatedness.

Most automatic image annotation techniques which are reviewed in the literature review part in chapter 2 used previously tagged images to recommend candidate tags for the new image using visual similarities and tags occurrences information. Some but not few used tags co-occurrence analysis of visual similar images with the query image to predict candidate tags for unseen image. In our case also we have used visual similar images' tags in addition to Wikipedia text documents to predict tags of a new image when it is uploaded without tag.

Chapter 4 Image Annotation in Folksonomies

4.1 Overview

In this chapter, we present the detailed description of the proposed approach which is Image Annotation in Folksonomies(social tagging systems, e.g. Flickr). First we provide the general Framework of the proposed approach, which is followed by the semi-automatic and automatic tags recommendation techniques with the metrics used in these methods. The measurements are used in finding tags dis(similarity) and tags feature selection.

In the process of semi-automatic image tagging, procedures to recommend cue-words for an ambiguous term(original tag supplied by the user) is proposed and if this tag is not suggested, a technique which automatically recommend tags for the new image by analyzing tags of previously tagged images is presented.

4.2 General Framework

In social tagging systems, since users are using freely chosen tags in an uncontrolled manner, the tags(or keywords) in these systems can face the problem of ambiguities. Different users may assign tags to documents in different ways-each user has subjective interpretation of her/his or others image content to provide annotation tags. Tags which have been attached as metadata to resources will decrease the performance of tagging systems regarding to search and retrieval of relevant items, since users contributed tags can be ambiguous (Abel et al., 2010).

Moreover, in these tagging systems, users may not provide tags for the resources(in our case to the images) they are uploading for some reasons. As a result of these problems, the organization, communication and retrieval of socially tagged web resources can not be optimal as required since images without tags can't be retrieved at all using textual information.

In this study we present a general framework to improve tagging in Folksonomies by integrating both tags refinement and propagation through analyzing tags relatedness. The framework and the flow of actions in the proposed approach are depicted in figures 4.1 and 4.2, respectively. In the figures, T_0 represents initial user provided tag set, T_1 represents the final recommended tags set and I_0 represents the new image uploaded by the user.

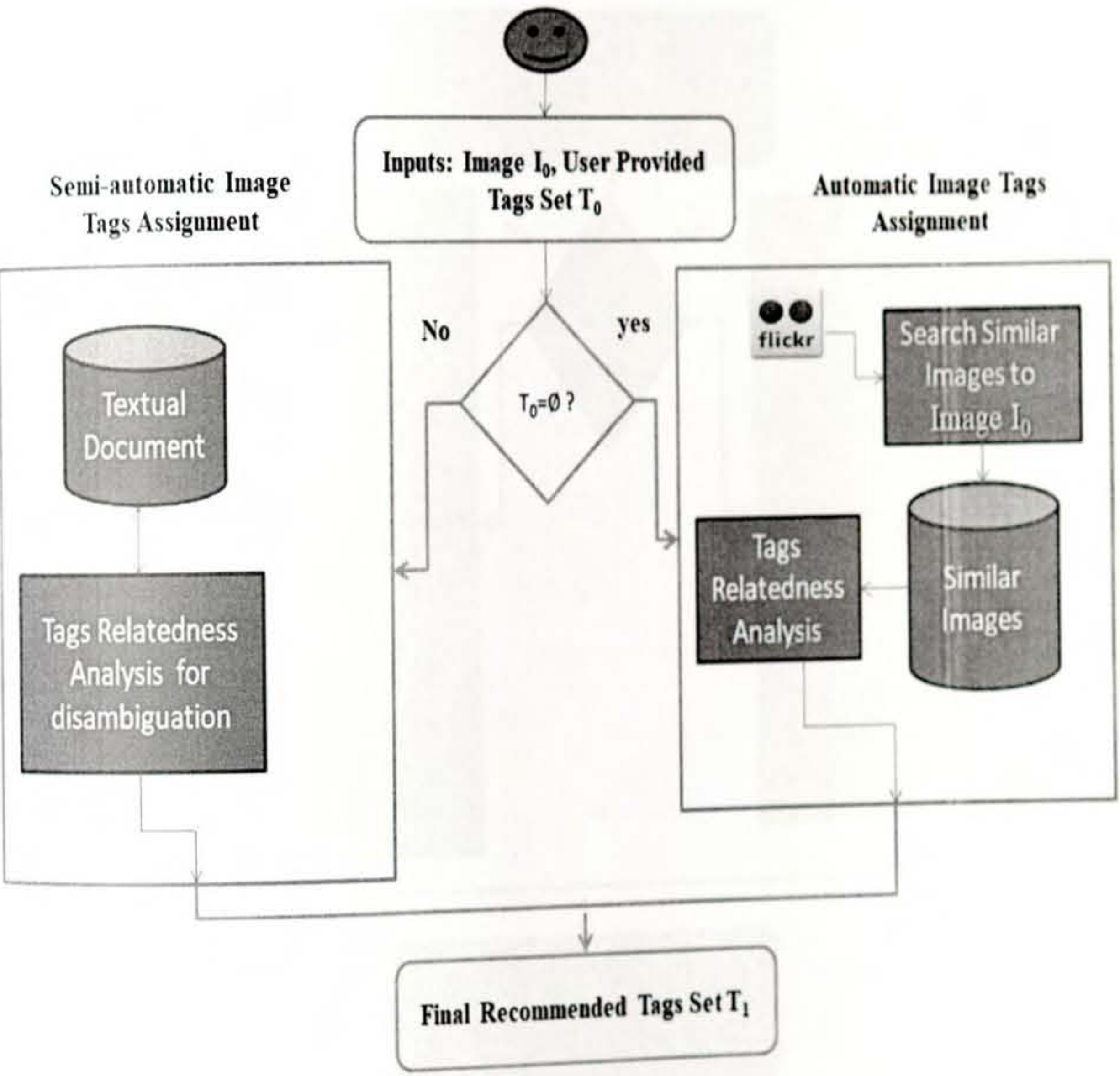

Fig. 4.1 General Framework of the Proposed Approach

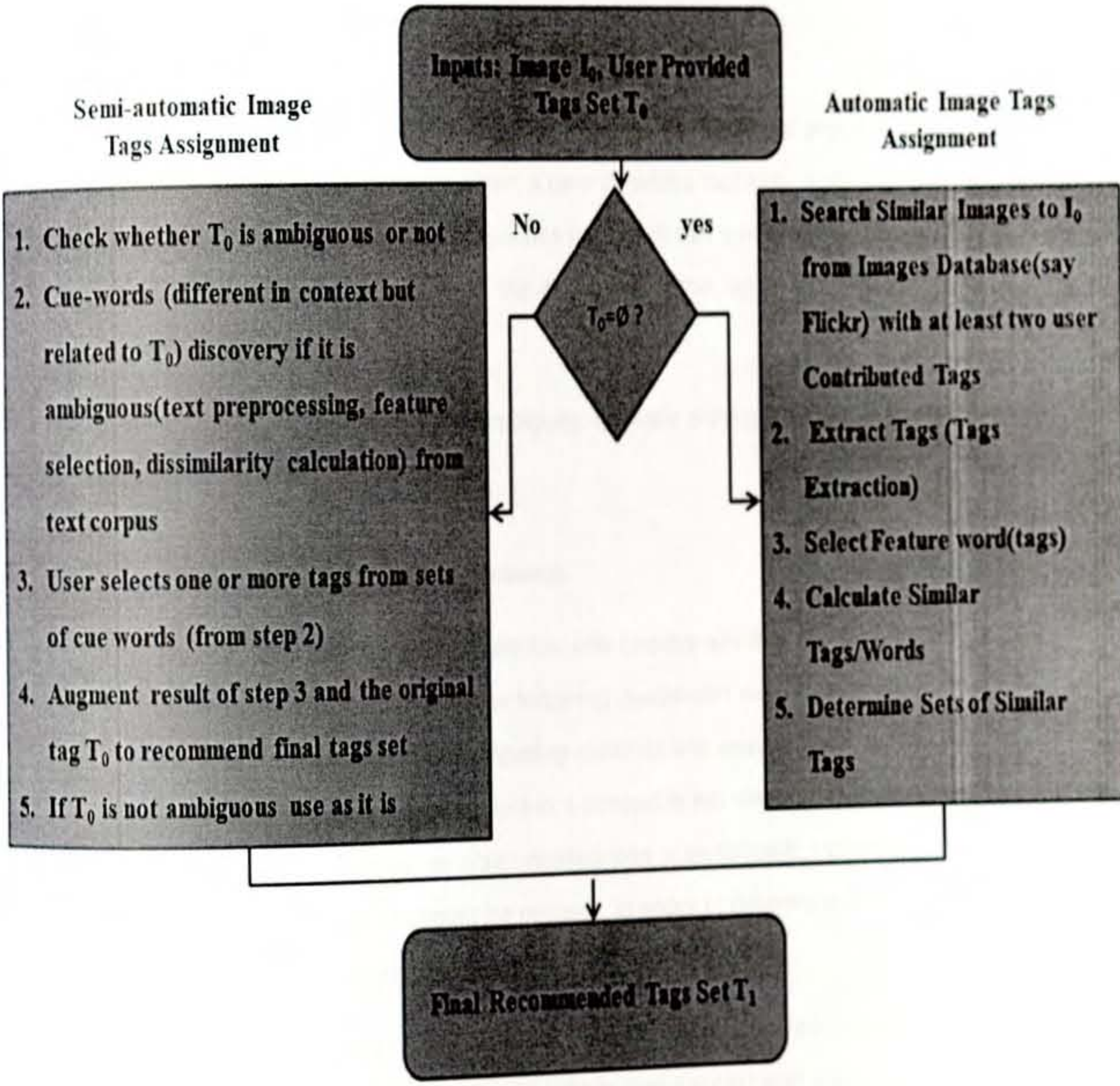

Fig. 4.2 Flow of actions in the Proposed Approach

4.3 Semi-automatic Image Tags Recommendation

4.3.1 Introduction

In this section, we proposed a procedure which is useful in the process of user tagging process. In the process when a user provides tag(s) to a certain image, the tag may be ambiguous. Our procedure will check this ambiguity and suggests possible tags to the user to resolve the ambiguity. if the tag is not ambiguous, it will be taken as it is.

The procedure, to resolve tag ambiguity, typically consists of in the following steps.

1. A user proposes a tag for a resource.
2. The tag is checked for ambiguity i.e. one checks whether the tag is *related* to more than one context (in the following discussion we will consider for the sake of simplicity that the competing contexts are exactly two). Notice that the simple appearance of a word in a context is not enough to assert word-the simple appearance of a word in a context is not enough to assert word-to-context relatedness: to claim relatedness a statistically significant word-to-context correlation should be present, in order to determine the quasi-co-occurrences of terms(or words/tags).
3. If the tag is found to be ambiguous, two cue words are sought, satisfying two requirements: each cue is specific to one context and each cue, within its context is related to the ambiguous word; in other words the two cue-tag pairs are related to the two contexts but are not ambiguous.

4. The user is provided with the list of cue words and chooses the one or more words which are associated to the meaning she/he has in mind for the original tag(s).

It is evident that the above procedure has to rely on effective word-to-context *relatedness* metrics. It turns out that, among the most effective relatedness metrics (see for instance (Begelman et al., 2006; Gemmell et al., 2008b; Papadopoulos et al., 2010)), are those based on a *feature-vector representation* of the words, defined by word co-occurrence. That is, given a language corpus structured in few-words sets – typically sentences,one can characterize a word by building the histogram (or empirical distribution) of its co-occurrences with each of the other words in the corpus (further details are given in Section 4.3.2). In this representation, the words of the corpus play the role of *features* and their frequency of co-occurrence with the target word plays the role of the *intensity* of a feature. By using this representation, one can quantify the dis(similarity) or the relatedness of two words by comparing their histograms: words represented by two different histograms will occur in distinct contexts. This can be determined by comparing the distance between the histograms of the two words.

It is also clear that different metrics of dissimilarity(or similarity) between histograms can lead to different results, in terms of word relatedness, and eventually can be more or less effective in the different phases of the disambiguation procedure, specifically in the key phases (2) and (3), i.e. ambiguous tag *detection* and *cue words discovery*.

In the task of *cue words discovery* (phase 3), and, using the co-occurrence-histogram representation of the words, the performance of three different dissimilarity(or similarity) metrics are studied. The first is the commonly used Symmetrized Kullback-Leibler divergence (SKLD) between the two histograms. The second is the Jensen-Shannon divergence (JSD) between the two histograms, and the third is a metrics introduced in (Mousselly-Sergieh et al., 2014) and consisting of a weighted sum of divergence terms, which represents a Maximum Likelihood estimator of the Jensen-Shannon divergence between the two histograms. By keeping the denomination adopted in (Mousselly-Sergieh et al., 2014), we call this metrics Adapted Jensen Shannon Divergence (AJSD). Its distinctive aspect is that, by definition, this metrics takes into account the statistical fluctuations present in the empirical probability distributions and is more robust with respect to statistical noise than the bare JSD of the two histograms.

In the cue-words discovery phase (3) – given a target word to be disambiguated – it is possible to proceed according to the unsupervised approach suggested in (Weinberger et al., 2008) for the construction of channels. The approach consists of finding the pair of words in the corpus which maximizes a disambiguation score: such score increases with the dissimilarity between the two target representations, conditional, respectively, to the first or the second word of the pair (further details in Section 4.3.2).

It has been shown that, used in the score definition, the AJSD metrics, performs better in terms of quality of the output, than the JSD metrics and the SKLD metrics for all the investigated settings. We also show that its comparative advantage is greater when the statistics is lower, i.e. when the corpus is smaller. In order to work under controlled conditions and to compare the preciseness of their performances, we carry on our study using a synthetic corpus.

Feature selection technique, using Shepley Value Analysis(further detailed is given in section 4.3.3, is applied to determine the marginal contribution of each word toward the disambiguation procedure.

4.3.2 Tags Disambiguation procedure

The objective of a tag disambiguation procedure can be considered as the prediction of tags(words) that users would like to select for a resource in their mind which will disambiguate and/or add supplementary information to provide more informative to the image they are uploading

Below, the proposed tags disambiguation procedure(algorithm) is presented.

We assume that a user is in the process of tagging a resource(particularly an image), and that she/he proposed a tag word t. The system detects the word as a word appearing in different contexts (*ambiguous word detection*) (e.g. t ="palm", which can refer to the palm tree or to the hand palm), and proposes a pair of cue words (e.g. (i, j) =("tree","hand")), the user chooses one; at this point the resource is annotated either with the pair (t, i) or with the pair (t, j).

One would like to present the user with a word pair endowed with a high discriminating power. To this aim we use co-occurrence histograms to characterize contexts and operate at two levels: at the level of histogram comparison we use, the AJSD divergence (see above), while at the preprocessing level Shapley Value analysis based feature selection is used. The latter technique helps in defining what words are used in the co-occurrence count, i.e. it helps efficiently defining the context.

Given a target word t, to be disambiguated, we consider pairs of cue words (i,j). Based on a set F of reference feature words (the features set $\mathscr{F} = \{f_1,\dots,f_m\}$), which represents our reference vocabulary, we build the frequency histogram $k(f)$ of the co-occurrence of the pair $\{t,i\}$: each sentence containing t and i and a word $f \in F$ will contribute to the histogram of absolute frequencies with a unit count in the channel f; similarly we derive the histogram $h(f)$ of the co-occurrence of the pair $\{t,j\}$; normalizing we get two empirical probability distributions.

The Adapted Jensen-Shannon Divergence is a divergence metrics similar to the standard Jensen-Shannon Divergence, but that takes into account the presence of statistical fluctuations in empirical probability distributions. Let us give the definition of the AJSD in line of our tags disambiguation procedure.

Let us consider the set T of possible tags and the set of possible feature words F. Let us consider the tag words $t,i,j \in T$: the first is the target word, the other two play the role of cue-words. Given a corpus (consisting in sentences or in tag sets used to annotate resources) the pair t, i co-occurs with other words of the feature set, the number of co-occurrences with the feature word$f \in F$ is denoted by $k(f) = \neq (f|t,i)$; similarly the co-occurrences with the pair{ t, j} determine $h(f) = \neq (f|t,j)$. We normalize and get the arrays

$$x_f = \frac{k_f}{n} \qquad y_f = \frac{h_f}{m}$$

representing two empirical probability distributions over P the same set of indexes $f \in F$, with $n = \Sigma_{f \in F} k(f)$ and $m = \Sigma_{f \in F} h(f)$

The Jensen-Shannon Divergence is defined by

$$d_{JS}(x,y) \equiv \frac{1}{2} \sum_f \left(x_f \ln \frac{2x_f}{x_f + y_f} + y_f \ln \frac{2y_f}{x_f + y_f} \right)$$

Let us define $z(f)$ as follows

$$z_f \equiv \left(x_f \ln \frac{2x_f}{x_f + y_f} + y_f \ln \frac{2y_f}{x_f + y_f} \right)$$

Let us take x_f and y_f as estimates of true quantities and we approximate their variances by

$$\sigma^2(x_f) = \frac{k_f}{n^2} \qquad \sigma^2(y_f) = \frac{h_f}{m^2}$$

then

$$\sigma^2(z_f) = \left(\ln \frac{2x_f}{x_f + y_f} \right)^2 \sigma^2(x_f) + \left(\ln \frac{2y_f}{x_f + y_f} \right)^2 \sigma^2(y_f)$$

now compute the weights $w_f = \left(\sigma^2(z_f)\right)^{-1}$.

The Adapted Jensen-Shannon Divergence is defined by

$$D_{AJSD(x\|y)} \equiv \frac{\sum_f w_f z_f}{\sum_f w_f} \tag{4.1}$$

Finally, based on the AJSD divergence of the two empirical probability distributions we compute a discriminating power score,a score used to quantify the discriminating power of a pair (i, j), with $i, j \in F$, in the disambiguation of a target word t, $DPScore$ (see below): all the other factors being equal, the higher the AJSD divergence – and consequently the score – the more different are the two distributions.

The intuition behind this is that very different distributions imply different contexts: a high difference between the distributions suggests that the cue words i and j are representative enough of their respective context to disambiguate the target word.

The procedure is stated as follows

1. Select, from a (possibly large) sentence collection (the corpus) S, the subset of sentences S_t containing t [1]

2. Text preprocessing on S_t

   (a) Remove stop words and punctuation

   (b) Get word stems using Porter Stemer the algorithm in ((Porter, 1980)) revised and published in (Willett, 2006) . This is important to avoid words redundancy.For instance, the words "run" and "running" will be considered as "run" .

3. Select the whole set V of unique terms (the vocabulary) in S_t

4. Find a set $F \subset V$ of $|F|$ representative terms (it will have the role of features, e.g. the set of $|F| = 100$ most frequent words in the corpus (not including stop-words))

5. (2-dimensional count histogram/array) For each term $v \in V$ and for each feature-word $f \in F$ count the number $c_2(v,f)$ of sentences in S_t in which v co-occurs with the word f, this will yield a 2-dimensional $|V| \times |F|$ count table

6. (1-dimensional count histogram/array) For each of the $|V|$ terms $v \in V$ [2] count number of entries in the v row $C(v) = \Sigma_f c(v,f)$ and use it for normalization: the probability associated to f given v will be

$$P_v(f) = c(f,v)/C(v)$$

this will yield a 1-dimensional array P_v (of length $|F|$) for each term $v \in V$

[1]Notice that the set that contains all possible form of t, i.e. all possible morphological form, should be determined.

[2]Notice that $V = F \cup \bar{F}$ and that v may belong either to F or to its complement $\bar{F}$: the most interesting terms for disambiguation, might come from $\bar{F}$

7. For each pair of terms $v, v' \in V \times V$, let us call $i = v$ and $j = v'$ compute a two-term divergence $d(i, j)$ (either the SKLD, the JSD or the AJSD) between the feature vectors $P = P_i$ and $Q = P_j$.

8. Compute the single-term probability $P(i)$ of a single term $i \in V$ by counting the number of sentences of S_t in which it occurs and dividing from the total number of sentences $|S_t|$ in S_t

9. Compute the disambiguation score of the pair (i, j) with respect to the word t as

$$DPScore_t(i, j) = P(i)P(j)g(D_{AJSD}(x_i, x_j)) \tag{4.2}$$

For example $g(d) = d^2$ or $= d$ or $= d^{\frac{1}{2}}$, while $P(i) = P(i|t)$ and $P(j) = P(j|t)$ represent the empirical probability of occurrence of the words i and j in sentences containing t. .

In the last expression d represents one of the above mentioned dissimilarity metrics: the SKLD, JSD or AJSD metrics defined in the previous section, with the following correspondence to the previous notation

$$k_f = c_v(f) \quad h_f = c'_v(f) \quad n = C(v) \quad m = C(v')$$

We have compared these three divergence metrics and proved that the AJSD measure is superior than the others. More discussion have been given in data analysis and experimentation chapter. The algorithm is stated as follows.

In the following subsection 4.3.3), describe the feature selection algorithm which is based on the Shapley Value.

Algorithm 1: Tags Disambiguation Algorithm (Cue-words discovery)

Data: t-term to be disambiguated, S(Corpus)-sentences collection
Result: DPScore of pairs of words with respect to t
Initialize S_t=$\emptyset$
$S_t = S$, if $t \in S$, select all sentences containing the term t and to S_t
Text preprocessing on S_t and assign terms of S_t to $S_{t'}$
$V = S_{t'}$, if $t' \notin V$, Vocabulary of unique terms
$F = mostFrequentTerm(V)$, Most frequent terms in V, say $|F| = 100$
for $\forall\, v \in V, \forall\, f \in F$ **do**
    $c(v,f)$=$counts(S_t)$,the number of sentences in which v and f co-occur in S_t and $C(v)$=$\sum_f c(v,f)$
end
Calculate the probability associated to f given v as: $P_v(f) = c(f,v)/C(v)$
$\forall\,(v,v') \in V \times V$, let us call $i = v$ and $j = v'$ compute a two-term divergence $d(i,j)$ (either the SKLD, the JSD or the AJSD) between the feature vectors $P = P_i$ and $Q = P_j$.
Compute the single-term probability $P(i)$ of a single term $i \in V$ by $P(i) = Count(i,s_t)/|S_t|, \forall i \in V$ count the number of sentence which contains i and divide by the total number of sentences($|S_t|$)
Compute the disambiguation score of the pair (i,j) with respect to the word t as $DPScore_t(i,j) = P(i)P(j)g(d_k(x_i,x_j))$, where k is either SKLD or JSD or AJSD

4.3.3 Feature-Words Selection Problem Based on Shapley Value

In order to characterize a context, we use word co-occurrence counting (see previous Section). Different contexts determine different count distributions, however not all the channels of the histograms contribute in a determinant way to the discrimination: if a word is irrelevant it will have more or less the same counts in all the contexts, furthermore if a word is redundant to another their counts will increase or decrease together when changing contexts.

This thesis work proposes a technique for feature-words selection to achieve high performance in word sense disambiguation and to determine co-occurrences. These can be seen as a classification process, thus the problem addressed consists in selecting a number k of features in a setting with a large number of features, so that the chosen set has the highest classification performance. Since each set of k features has a classification performance that is not known a priori, the problem corresponds to a search problem over the Boolean lattice of the k-subsets, of a given base set of size n. Since the classification performance function of real settings in general does not offer any guaranties of nice behavior, which could simplify the search (e.g. monotonicity) the search problem corresponds to the unconstrained search problem and is known to be Non-deterministic Polynomial hard(NP-hard).

In this work we use an analogy with coalitional Game Theory and see the possible feature sets as coalitions: we propose a heuristic based on the Shapley Value of the n set elements (the Shapley Value captures the average marginal contribution of an element to all the possible coalitions). We exploit the intuition that there is a strong correlation between having the top-k Shapley Value and the membership of the k-subset with top classification performance. We use the heuristics consisting in using the top k Shapley value elements to select the k features of the candidate feature coalition.

We prove by simulation that the heuristics is more effective than the traditionally used naïve approach consisting in selecting the features only on the base of their individual classification performance, measured in isolation. The Shapley Value-based approach is more effective because it takes into account also the interaction (e.g. redundancy) among features.

The frame for feature selection methods are "filter" methods which select the subset making no explicit reference to an algorithm/task (e.g. a classifier), but on the basis of the properties of the features themselves, and "wrapper" methods which select the subset with the help of an algorithm used as a black-box to evaluate the quality of a subset.

Our algorithms belong to the wrapper category since we use a classifier as a black-box to evaluate the performance of the feature sets. This technique is used to evaluate a procedure how good it is at producing quality or desirable answers(outputs) (Nyberg and Mitamura, 2002).

Naive wrapper methods rank the features performance evaluated individually and then select the top k performing elements: such naïve methods would make perfect sense in additive settings, which is not the general case, due to synergies and redundancy.

Greedy wrapper methods are the forward-selection algorithm and backward-elimination algorithm (on adds one element at a time, in forward selection one adds the element best performing with the current coalition, in backward elimination one drops the worst element from the current coalition). Slightly improved variants of those greedy methods are best first-search (less greedy, it does some revisiting) and beam search (which restricts the search to subsets of promising candidates). Greedy wrapper methods make perfect sense in monotonic settings. However, in the real word cases the addition of features is not monotonic in the classification performance, due to interaction among features.Greedy wrapper methods work very fast, but find local maxima.

The rationale behind our approach is that those players, out of the n players, that individually interact the worst on average are more likely not to be in the best-performing coalition of size k, thus choosing the top k Shapley Value players we approximate the best coalition of k elements. In other words, we exploit the correlation between having the top-k Shapley Value and the membership of the k-subset with top classification performance. This heuristic does not rely on any monotonicity or super-modularity hypothesis. The results of the simulation support the soundness of this approach.

Let V a full vocabulary of $n = |V|$ terms. Let $F \subset V$ a subset (it will contain selected feature words) of size $k = |F|$.

The set V has a performance measure $\mu(V)$. Such performance measure depends on the formulation of the problem: it could be for instance precision, or recall or F1-score. For the sake of convenience we use $\mu(V)$ as a reference and normalize it to 1 so that: $\mu(V) = 1$. The performance measure of F is $\mu(F)$.

Any set F has also a cost $c(F)$ that we assume monotonically increasing with its size, so that $c(F) = c(|F|) = c(k)$. This cost might represent the computational cost for computing the features.

Our problem feature selection problem is the following: we would like to select a subset $F \subset V$ with $k << n$, so as to reduce the cost (for monotonicity $c(k) < c(n)$), while at the same time we safeguard or improve the performance. We would be satisfied even if there is some limited loss of performance, in view of the fact that we reduce the costs. We wish $\mu(F) \geq \mu(V) - \delta$. Finding the *optimal* set corresponds thus is finding F such that

$$\mu(F) \geq \mu(F') \quad \forall F' \in 2^V$$

$$\mu(F) \;\geq\; \mu(V) - \delta$$

The key point is that the performance function is *non-additive*. Should it be additive (denote the elements of $V \equiv \{1,2,\ldots,n\}$), one coud compute the performance of the features in isolation $(\mu(1),\mu(2)\ldots)$ then sort them and select the top k features: this would solve the problem. This scenario corresponds to a setting where the features do not interact: e.g. in a detector/classifier selection problem, each feature provides an independent classifier because one focuses on different parts of the data space.

If the performance μ is non-additive, the problem is more involved and requires combinatorial optimization.

Obviously, a brute force exhaustive search has typically a very large computational cost. Let us indicate by q^{EX} the cost of the exhaustive search. Checking all the sets of size k will correspond to

$$q^{EX} \propto \binom{n}{k} \underset{n\to\infty}{\to} n^k$$

where $\underset{n\to\infty}{\to}$ represents the large n limit.

Nonetheless, we can simplify the search for a solution if we can assume that the set has some structure.

For instance if we could assume that the features, although not additive, interact only in pairs, and so on. This would limit the search to the evaluation of all the individuals and all the pairs.

However, we are not in condition to make a priori assumptions of this kind and would like to adopt a general-purpose approach.

One such general-purpose approach uses the evaluation of the "potential of collaboration" of which each individual feature is endowed, and then the selection of the k "top-potential" features.

We assume that the player has a potential value that he/she can express within a collaboration. This is quantified by a power index: we choose as power index the Shapley Value.

We estimate the Shapley Value of all the features (their importance in improving the performance metrics, e.g. F-score), then we sort them by their Shapley Value and select the top k.

Our experimental work consists in verifying that sorting them by Shapley Value and taking the top k produces feature sets that have higher performance than those that would be obtained by sorting the features by F-score and keeping the top k.

The metrics $v(C)$ which quantifies the value of a set of feature words in our case is the objective function of our problem: the capability of providing effective feature words, measured in terms of a chosen metrics-the AJSD. As we have proved such metric is defined as the precision of a restricted set of 20 highest score disambiguation pairs[3].

[3]Notice that 20 is selected arbitrary by observing the consistence of the result using it other than numbers below it. When numbers above it are used the precision score starts decreasing

A correct disambiguation pair is such that the two words belong to two distinct contexts, an incorrect disambiguation pair is composed, on the contrary, by words from the same context. Any candidate set of selected feature-words C will be evaluated through a function defined as the rate of correct pairs in the top scoring 20 pairs:

$$v(C) = \frac{1}{20}(\#\textit{correct top-score pairs when } C \textit{ is the feature set}) \tag{4.3}$$

Consider a term t to be disambiguated, we use the following *algorithm*.

1. Select, from a (large) sentence collection (the corpus) S, the subset of sentences S_t containing t [4]
2. Text preprocessing on S_t
   (a) Remove stop words and punctuation
   (b) Get word stems using Porter Stemer the algorithm in ((Porter, 1980)) revised and published in (Willett, 2006).
3. Select the whole set V of unique terms (the vocabulary) in S_t
4. Find the set $F \subset V$ of $|F|$ representative terms, i.e. of selected features, using the highest Shapley Value features as follows
   (a) Initialize the weights of all the features $f \in V$ to zero; i.e. let the weight of feature be indicated by $B(f)$: set $B(f) = 0$ for all $f \in V$

[4] Notice that the set that contains all possible form of t, i.e. all possible morphological form, should be determined.

(b) Compute the Shapley Value $SV_{(v)}(f)$ of a feature word.

Generate a large number of permutations of features and for every permutation π

i. initialize an empty set C of candidate features: the evaluation of this set is 0 by definition (i.e. $v(C=\emptyset)=0$))

ii. for every element f of the permutation π

A. add f to the current set C of features obtaining $C \cup f$ and call the evaluation function $v(C \cup f)$ on the new set (evaluate the current performance or importance of a feature)

B. take the difference between the old and the new value: $\Delta(f) = v(C \cup f) - v(C)$

C. add the value to the balance of the element f

(c) normalize the balances $B(f)$, i.e. compute the sum of the $B(f)$ and divide each $B(f)$ by this total: the normalized value is the Shapley Value of the feature.

5. Select the 100 feature words with highest Shapley Value as features set.

We have used both synthetic and real world corpora to test our algorithm and more precisely the procedure is defined in Algorithm 2.

Algorithm 2: Shapley Value Based Feature Selection Algorithm

Data: t-term to be disambiguated, S(Corpus)-sentences collection

Result: Set of feature words with corresponding weighted Shapley Values

1 Initialize S_t=$\emptyset$

2 $S_t = S$, if $t \in S$, select all sentences containing the term t and to S_t

3 Text preprocessing on S_t and assign terms of S_t to $S_{t'}$

4 $V = S_{t'}$, if $t' \notin V$, Vocabulary of unique terms

5 Find the set $F \subset V$ of $|F|$ representative terms, i.e. of selected features, using the highest Shapley Value features as follow

1. Initialize the weights of all the features $f \in V$ to zero; i.e. let the weight of feature be indicated by $B(f)$: set $B(f) = 0$ for all $f \in V$
2. Generate a large number of permutations of features to Compute the Shapley Value $SV_{(v)}(f)$ of a feature word.
3. **for** *every permutation* π **do**
   - initialize an empty set C of candidate features: the evaluation of this set is 0 by definition (i.e. $v(C = \emptyset) = 0)$)

   for *for every element f of the permutation* π **do**
   - add f to the current set C of features obtaining $C \cup f$ and call the evaluation function $v(C \cup f)$ on the new set (evaluate the current performance or importance of a feature)
   - take the difference between the old and the new value: $\Delta(f) = v(C \cup f) - v(C)$
   - add the value to the balance of the element f: $B(f) = B(f) + \Delta(f)$

   end

   end
4. normalize the balances $B(f)$, i.e. compute the sum of the $B(f)$ and divide each $B(f)$ by this total: the normalized value is the Shapley Value of the feature: $B(f)$=$B(f)/\Sigma_f B(f)$

6 Select the 100 feature words(for instance) with highest Shapley Value as features set.

4.4 Automatic Image Tags Recommendation

4.4.1 Introduction

The aim of this section is proposing an automatic technique to generate tags for the image contributed in social tagging system. The procedure(algorithm) is triggered , if an image is uploaded without initial tag.

The user may be in a hurry or may not know the content of the image to provide tag(s). In this case, similar images from the web will be searched and their tags will be assigned as annotation to the new image. To minimize the aforementioned tags' problems (scarcity and ambiguity) we exploited the importance of tags co-occurrences. Here we have used the AJSD measure to find the divergence between two tags'/words' histograms and the Shapley Value Calculator to measure the average marginal contribution of each tag in the corpora (obtained from previously tagged images and from Wikipedia text document) toward finding co-occurring tags.

4.4.2 Tags Generation Procedure

Let us assume a user provided an image without tag(s). We have developed a procedure which can generate possible tag(s) for the untagged image automatically. The procedure is stated as follows.

1. Search similar images with at least two user contributed tags(pair of tags has more semantic than a single tag) from the web(Say Flickr image database)

2. Determine frequent tags set, let us call it $T = \{t_1, t_2, ..t_n\}$ (for example n=10: this will be determined by taking the average number of tags per images of similar images)

3. Select from the sentence collection (the corpus, each image tags considered as words in a given sentence-for the corpus obtained from previously tagged images or sentences in text document for the Wikipedia text document) S, the subset of sentences S_t containing either element of T

4. Text preprocessing on S_t

   (a) Remove stop words and punctuation

   (b) Get word stems using Porter Stemer the algorithm in ((Porter, 1980)) revised and published in (Willett, 2006)

5. Select the whole set V of unique terms (the vocabulary) in S_t

6. Find the set $F \subseteq V$ of $|F|$ representative terms, i.e. of selected features, using the highest Shapley Value of features.

7. Determine Conditional probability of each tag given the tags set T i.e; $p(t_i) = \frac{n(i,T)}{n(T)}$, where, $n(i,T)$ is number of times i co-occur with T and $n(T)$ number of elements in T

8. Determine Co-occurrence Score of t_i and t_j using

$$CoScore(t_i, t_j) = P(t_i)P(t_j)g(d(t_i, t_j)) \tag{4.4}$$

where, $d(t_i, t_j)$ is the divergence using the Adapted Jensen-Shannon Divergence stated in the formula 4.1 , which is better than the other known divergence metrics and same tuning function of subsection 4.3.2

The aim of the search step in the algorithm is to find images(with tags) which are similar to that in the target (or query) image. Content-based image retrieval is a well-studied paradigm and we wouldn't have interest in improving searching techniques to search similar images to the target image. So, we have used the Flickr content-based image retrieval(An application program interface (API)) to search similar images. This is to analyze and use the tags of these similar images as tags of the images uploaded without tags. The similarity search by Flickr API uses deep neural networks for deep understanding of image content[5].

To implement our procedure we have represented the relationships among images and tags by a matrix. In the matrix each row corresponds to an image and each column corresponds to a tag. We have used the following function (ImageTag) to determine the corresponding numeric value for each cell in the matrix:

$$ImageTag(i,j) = \begin{cases} 1, if \ and \ only \ if \ image \ I_j \ annotated \ by \ tag \ t_i \\ 0, otherwise \end{cases} \tag{4.5}$$

[5]https://yahooresearch.tumblr.com/post/158115871236/introducing-similarity-search-at-flickr, accessed on September 5, 2018

Then the matrix is transformed to Tag-Tag matrix-which is tags Co-occurrences matrix formed using the following function(TagTag):

$$TagTag(t_i,t_j) = \begin{cases} n, the\ number\ of\ images\ where\ tags\ t_i\ and\ t_j \\ simultaneously\ co-occur\ as\ annotations \\ 0, otherwise \end{cases} \tag{4.6}$$

From the set of Users, Tags, and Images(Resources) of a *Folksonomy*, let us consider the associations between Tags and Images only. Let us represent these associations as matrix $A = a_{ij}$, where $a_{ij} = 1$, if and only if tag t_i is used to label image I_j and 0 , otherwise as mentioned in equation 4.5. Based on this notion we can determine the Tag-Tag matrix. Let us denote this matrix by B, as it has been shown by equation 4.6 and B is given by:

$$B = AA' \tag{4.7}$$

Where A' is the transposed matrix of marix A and element b_{ij} of B(Tag-Tag matrix) represents the number of images in which tags t_i and t_j appear together as annotations.

From this notation, B is a Tag-Tag similarity matrix since it represents the co-occurrences of tags. These similarity are from the absolute co-occurrences of tags, because it only counts the number of times two tags co-occur as tags of a given image. To determine similarities beyond absolute similarities, we have used the proposed similarity measure the Adapted Jensen-Shannon Divergence measure as indicated in the algorithm in this section.

We have used a Flickr API to collect similar images to the target image. Then we have built tags co-occurrences matrix to find similar tags. In our context by similar we mean tags co-occur as annotation of the same images. Let i and j are two tags, we say t_i and t_j are similar if they are annotations(or tags) of a certain image W.

From the image-tag and tag-tag matrices defined in formulas 4.5 and 4.6, we can find the co-occurrence of tags t_i and t_k by finding the number of images that both occur simultaneously in many images as tag. In such ways if t_i is associated with w_j, we can recommend t_k to w_j . The underlined principle of co-occurrences based methods is similar to that of collaborative filtering (Adomavicius and Tuzhilin, 2005).

We have used feature selection algorithm defined in (subsection 4.3.3) to select important tags(features) which contribute most (to optimize our results) toward finding of tag co-occurrences. In the procedure, the correct pair indicates the two words co-occur as annotation of the same image. Any candidate set of selected feature-tags C will be evaluated through a function defined as the rate of number of correct pairs in the upper 20 pairs obtained from previously tagged images. We have used equation 4.3 for the evaluation purpose.

After features set is determined, the final set of tags would be generated from the upper 20 pairs as annotation of the new image which has been uploaded by the user without tag(s).

The procedure has been tested using corpora from tags of previously user provided tags of images and textual documents from Wikipedia.Detailed discussion found is in the data analysis and experimentation chapter.

4.5 Summary of the Chapter

In this chapter, we have presented the proposed solution to annotate images in social tagging systems. In the proposed solution we have two cases:

1. A user can upload an image with initial tag, or
2. A user can upload an image without tag

If a user uploads an image with tag, we have proposed a procedure which will check the tag to see whether it is ambiguous or not and if it is found to be ambiguous, list of cue-words which help to disambiguate the original tag will be recommended. The user will select one or more words from the recommended cue-words list that matches the one in his/her mind which can resolve the ambiguity and this will be augmented with the original tag(initial tag) as a final annotation for the new image. If it is not ambiguous, the tag will be taken as it is as annotation of the new image

The second case is when the user uploads image without tag. In this case we proposed a procedure to recommend tags for the new image from previously annotated similar images automatically.

In both cases we have used AJSD measure to determine the distance between two probability distributions in finding similarity and dissimilarity of tags and features selection technique to select important words/tags which improve the performances of our procedures.

Chapter 5 Experimentations

5.1 Overview

In this chapter, we describe the experimental setup for analyzing the performance of the proposed approaches, including the data sets used, the testing procedure, and the evaluation metrics. We have used both synthetic and real world datasets to evaluate our proposed approaches.

5.2 Experiment Goals

Main goals of experimentations in this thesis are to check the appropriateness of tags relatedness metrics in determining quasi co-occurrence(true co-occurrences) of tags and select the best one, and to evaluate Shapley Value based feature selection strategy in evaluating the importance of each feature(i.e. a tag) toward determining of word relatedness.

Different probability distributions divergence measures used to determine the distance (divergence) between probability distribution histograms are compared in finding dis (similarities) of tags/words. Divergence metrics, such as SKLD, JSD and AJSD, were experimentally learned to determine tags relatedness in this experiment. The second goal is to demonstrate the validity and performance of FS based on Shapley Value as compared to the state of the art with Feature

Selection(FS) based on Most Frequently Occuring Words(FOWs) on the same experimental setup.

Generally, we have showed that the extent to which each of the modifications in our approach contributes to the overall performance improvements: the first one is in cue-tags/words determination in tag disambiguation procedure for a typical semi-automatic image annotation and the second one is in automatic image annotation procedure to determine set of tags based on tags co-occurrences.

5.3 Comparing the Dis(similarity) Metrics

5.3.1 The dataset to Compare Dis(similarity Metrics)

In order to compare the dis (similarity) metrics toward the performance of the proposed tag relatedness strategy in the disambiguation process, experiment on synthetic corpus is performed. To carry on our study under controlled conditions, we opted to create a synthetic corpus. Using this corpus we have compared the three dis(similarity) metrics(SKLD, JSD, and AJSD) defined in section 2.3 with the procedure in subsection 4.3.2.

- We created two distinct vocabularies of 1000 words each, representing the vocabulary of two distinct sub-domains, which we call here domain A and domain B. To domain A vocabulary, we associated conventional id's from 1000 to 1999, and to the domain B vocabulary,we associated the id's from 2000 to 2999.

- Then for each domain we generated a large number of 3-word sentences, picking the words within each domain (the number of sentences was a parameter of the experiment) . Each sentence was composed only by words of the same domain,[1] using an index-difference based randomization procedure: the latter convention was motivated by the choice of providing some structure to the domain, so that the probability of co-occurrence of any word with any other word in the domain was not uniform. The random generation would proceed as follows:

  1. A domain would be chosen at random between A and B.
  2. The first word of the sentence would be generated uniformly at random within the corpus,
  3. The second word would be generated at random according to a normal distribution, centered on the index of the first chosen word, and
  4. The third word would be generated at random according to a normal distribution centered on the index of second chosen word. For the standard deviation of the Gaussian we chose the value 10. No cross-domain sentences were allowed.

- After this we chose a fixed word from the domain A, say the "word" "1111" and a fixed word from the domain B, say the "word" "2222" and replaced those two words in all the sentences of the corpus with a third word, say "3333", in this way belonging neither exclusive to A nor to B; this word would play the role of the word to disambiguate.

[1] The order of the words in the sentences did not count, since our procedure treats a sentence as a bag of words.

5.3.2 Experiment in Comparing Dis(similarity) Metrics

After the generation procedure of words, the corpus consists in a large number of sentences, whose domain of origin is always clear. Those sentences are not to be used as target sentences for word disambiguation, but rather as reference sentences to be used as background knowledge.

At this point we assume that a hypothetical user is in the process of tagging a hypothetical resource with the tag "3333": we assume that the tag is recognized as ambiguous, since it is significantly present in both domains, the task now consist in proposing to the user a pair of words so that she can choose one she deems associated to "3333" so as to resolve this word ambiguity.

The task is accomplished by assigning a score – computed by the procedure illustrated in Section 4.3 and by equation (4.2) – to every pair of words possible according to the vocabulary. The full vocabulary consists of $n = 2000$ words, hence the possible pairs are $n(n-1)/2 \simeq 2 \times 10^6$. Notice that this includes both same-domain, and different-domain pairs.

Ideally, if the procedure is successful, same-domain pairs should be given a low disambiguation score, different-domain pairs a high score. However the statistical noise interferes with the correctness of the procedure, so that – especially for smaller corpora – one can find same-domain pairs also in the top-score region of the grading.

So as to take into account this problem, we took for each realization of the experiment (different in terms of corpus size) into consideration the top 20 candidate pairs and counted how many belonged to the different-domain category and results showed the procedure with AJSD divergence metric performed best.

5.4 Shapley Value Based Feature Selection

5.4.1 The dataset for Feature Selection

To evaluate our procedure stated in subsection 4.3.3 of section 4.3,which is Shapley Value based feature selection to determine cue-words in tag-sense disambiguation procedure, we have used the same dataset and size used in subsection 5.3.1 but using the Adapted Jensen-Shannon Divergence measure which has shown superior results than the other two(see section 6.2).

Besides, to evaluate the proposed procedure using a real world dataset, we have used a relatively large text document from *Wikipedia* English texts dump. We have chosen Wikipedia as source of textual documents, since it offers a much wider vocabulary than other readily available resources, for instance WordNet. From this dump, a 6+ GB document is downloaded and a Plain Text Corpus is generated using the technique for Text Extractor[2].

[2]https://blog.afterthedeadline.com/2009/12/04/generating-a-plain-text-corpus-from-wikipedia/, by rsmudge, 2009

5.4.2 Experiment on Feature Selection

To compare the performance of the procedure(in determining cue-tags/words in the disambiguation procedure) with two setups:(1) feature selection by most frequently occurring words, and (2) feature selection by top Shapley Value words both synthetic and real word linguistic datasets are used. Here, AJSD divergence measure that showed superior performance than SKLD and JSD in determining tags relatedness is used.

- For the synthetic dataset, the same setup of the above experiment(section 5.3.2) is used.
- For the real world dataset, corpus prepared from Wikipedia text document is used. From these text documents we have selected sample tags manually for performance evaluation. We assumed that each tag is provided by the user as a tag of an image uploaded by him/her or other in the folksonomy.

  Sample results obtained on the real world data set is presented in table 5.1. Let us assume the user provided a tag "Cambridge" for an image and for another one "Bank" , cue-words pairs and unique words recommended by the proposed algorithms 1 and 2 are shown on the second and last columns of Table 5.1, respectively. The last column presented list of unique tags recommended to the user.

Table 5.1 Sample output for the real world linguistic corpus

Target Word	Pairs of Words	Unique Words in the Pairs
Cambridge	(hotel,university),(hotel , tour),(hotel , city), (hotel , hospital) ,(hotel , bridge), (hotel , travel), (tour, university), (tour, city), (tour, hospital),(tour, bridge), (tour, travel), (university, city),(university, hospital), (university, bridge), (university, travel), (city, hospital), (city,bridge),(city, travel),(bridge, travel)	hotel, tour, university, city, hospital, bridge, travel
Bank	(coastal, deposit), (coastal, bond), (coastal, snow), (coastal, rate), (coastal, balance), (coastal, store),(deposit, bond), (deposit, snow), (deposit, rate), (deposit, balance), (deposit, store),(bond, snow), (bond, rate), (bond, balance), (bond, store), (snow, rate), (snow, balance), (snow, store), (rate, balance), (rate, store), (balance, store)	coastal, deposit, bond, snow, rate, balance, store

5.4.3 Evaluation Metrics

To quantify the results of our approach in determining tags relatedness in the disambiguation procedure with feature selections, standard information retrieval measures are used. These are precision, Recall and F1-Score(Goutte and Gaussier, 2005; Peddi et al., 2010; Pino, 1999; Powers, 2011; Robertson, 2000).

In the technique of proposing possible set of senses; top n pairs of tags are determined and the following measures are applied for n is equal to 20:

- Precision: measures the rate of correct pairs in the top scoring 20 tags pairs. A correct disambiguation pair is such that the two words belong to two distinct contexts, an incorrect disambiguation pair is composed, on the contrary, by words from the same context. Precision is computed as follows:

$$Precision = \frac{cTag20P}{Tag20P} \tag{5.1}$$

- Recall: measures the rate of correct pairs in all the resulting tags pairs. Precision is computed as follows:

$$Recall = \frac{cTag20P}{TagallP} \tag{5.2}$$

- F1-Score the harmonic mean of precision and recall. F1-Score is computed as follows:

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision} \tag{5.3}$$

Where, cTag20 is the number of correct tags pairs in the upper 20 pairs, Tag20P is the upper 20 tags pairs and TagallP is the number of all tags pairs in the resulting output and the results of the evaluation using F1-score is about 92% in accuracy using Shapley Value based feature selection which showed better performance than FS based on most FOWs.

5.5 Automatic Image Tags generation

Let us assume a scenario for the target users of the Flickr Web site: as members of the community or personal users, want to index (and share) latest pictures and upload them. In these processes, let us assume the user knows some of the images(or photos) they have and do not know others exactly or he/she may have not time to annotate. Tags may be assigned for those which have been known and others can be uploaded without tags.We want to generate tag(s) of the new image automatically for those images which have been uploaded without tags.

To address this problem we followed the procedure stated in section 4.4.2. After collecting similar images to the target image, candidate tags for the untagged image have been selected from the tags of visually similar images. We use tags that occur multiple times as tags of many images and tags relationships to other tags (using tags relatedness measure to exploit tags co-occurrence information) in this set of tags from similar images. In the divergence measure if pair of tags' divergence measure score is relatively less than those of other pairs, the probability of the co-occurrence of these tags as annotation of the same image is high.

5.5.1 The dataset and experiment for Generation of Tags Automatically

we conduct experiments on a dataset collected from a real-world system, namely Flickr, for online photo sharing site. The dataset is collected using sample images as user provided inputs, by assuming these images have been uploaded without tags.

We have collected 100 sample images and for each of them, similar images, with at least two user provided tags, are searched using the content of the input image. A corpus is prepared from these images.

The sample images are distributed among 10 human annotators' who have been trained toward the purpose of our procedure. Based on the training 2 annotators seated together and did the annotation. Their annotation results are checked and discussed to avoid subjectivity and approved for the test.

Additionally, we have used text corpus prepared in section 5.4.2 to test the performance of the procedure with respect to the semantic knowledge repository-Wikipedia. To make our text corpus relatively exhaustive, we have included some text documents searched from Wikipedia using keywords obtained from previously tagged images(step 2 in the procedure under section 4.4.2).

The sample results are demonstrated in the table 5.2. In this table tags are separated by commas and sorted alphabetically.

Table 5.2 sample Tags Results obtained from Human annotators and Our procedure

Sample Images					
Human Tagging	animal, big, cat, jaguar	Ethiopia, King, Lalibela	bete, church ,ethiopia, giorgious, lalibela	ethiopia, city, lalibela	beach, island,set, sun, tree
Our Procedure Tagging	animal, big, cat, jaguar,wild	church, Ethiopia, King, Lalibela	bete, church ,giorgious, hewn, lalibela, rock	lalibela church king city ethiopia	beach, island,set, sun

5.5.2 Evaluation Metrics of Generation of Tags Automatically

To quantify the performance of our procedure, which have been implicitly tested in the tag sense disambiguation under semi-automatic image tags recommendation procedure, we have used Precision, Recall and F1-Score effectiveness measures.

These performance measures are defined as follows:

Let HumanTag be the set of all tags provided by human annotators and ProcedureTag be the set of all tags recommended by our procedure from tags obtained from previously tagged images. Then

$$Precision = \frac{|HumanTag \bigcap ProcedureTag|}{|ProcedureTag|} \tag{5.4}$$

$$Recall = \frac{|HumanTag \bigcap ProcedureTag|}{|HumanTag|} \tag{5.5}$$

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision} \tag{5.6}$$

The results of the evaluation using F1-score is about 92% in accuracy.

5.6 Summary of the Chapter

In this chapter we presented experimentations to evaluate the performance of our proposed approaches using both synthetic(controlled vocabulary) dataset and real world datasets from Wikipedia text document dump and Flickr Images database.

The corpus from synthetic dataset is used to evaluate the preciseness of the metrics used in our procedures. Based on the results obtained from the synthetic dataset, we have extended our experiments on real world datasets.

To implement our algorithms, all the programming was done in Python . Python Libraries, packages, codes for different functions and modules are used from freely available source on the web.

Chapter 6 Discussion of Results

6.1 Overview

A Hybrid Image Annotation mechanism in Folksonomies is proposed using tags/words Co-occurrences (chapter 4). The proposed approach is believed to improve collaborative tagging whereby people attach tags to images, in order to render them retrievable in the future.

Improving the quality of tags in a typical semi-automatic image annotation environment is investigated and the possibility of automatically Identifying the semantically related tag pairs and making explicit their relationship, even in the absence of user provided initial tags is studied based on the research questions raised.

Hence, the purpose of this chapter is describing results obtained from experimentations, in order to show the extent to which the research questions are addressed. Besides, important findings from this study are presented.

Results on cue-words discovery in the disambiguation algorithm to compare different divergence measures(measures to determine the distance between two probability distribution histograms), on the comparison of cue-words discovery algorithm with feature selection based on frequent occurring words versus feature selection based on Shapley Value and on results obtained by best results of these techniques recommending tags of images automatically are discussed.

Results obtained are compared and demonstrated using diagrams and standard evaluation metrics.

6.2 Techniques of cue-words discovery for tag-sense disambiguation

6.2.1 Comparing Dis(similarity Metrics)

The various divergence measures to measure the distance between two probability distribution histograms and weighting techniques have been employed to increase the accuracy of tags, in a typical semi-automatic image annotation technique, based on co-occurrences for the algorithm we proposed to find cue-words in the disambiguation process is depicted in figure 6.1.

First, we want to discuss the comparison of the performance of the cue-word/tag discovery process in tags disambiguation procedure proposed in this study by applying different divergence metrics using sentences from synthetic corpus.The comparison is explained based on the same tag representation. This representation is based on a target word and its senses(or context). Secondly, we will compare the results after applying feature selection techniques in the next subsection 5.4.1 under this section.

The results referring to the corpus sizes of $10^2, 10^3, 10^4, 10^5$ and to two different choices of the stretching function $g()$ in equation (4.2) are shown in Figure 6.1. In each graph, the three divergence metrics SKL, JSD and AJSD are accounted for.

One can observe(see figure caption for details) that on average the number of correct pairs discovered by the AJSD metrics is always higher than the other two metrics, which turn out to be more or less equivalent; furthermore the advantage of AJSD is more apparent at low statistics, ie. with small corpora. We notice also, incidentally, that the function $g(d) = d^{\frac{1}{2}}$ has a better performance than $g(d) = d$.

From figure 6.1, labels on the x axis represent the logarithm in base 10 of the size of the corpus. The y axis represent the average number of correct disambiguation pairs discovered out of the 20 top-score pairs, using the score of equation (4.2) with, a) in the upper figure $g(d) = d$ and b) in the lower figure $g(d) = d^{\frac{1}{2}}$. One can observe that in average the number of correct pairs discovered by the AJSD metrics is always higher than the other two metrics, which turn out to be more or less equivalent. Furthermore, the advantage of AJSD is more apparent at low statistics, i.e. with small corpora, which mimic the typical real word situation in which the linguistic data are relatively scarce.

6.2.2 Shapley Value Based Feature Selection

In this section we want to discuss the comparison of our approach, Shapley Value based feature selection, with feature selection using most frequently feature which is the practice of most related work. For instance, most related works to ours, discussed in the related works chapter, indicated in (Sigurbjörnsson and Van Zwol, 2008; Weinberger et al., 2008) used FS based on most FOWS. By using AJSD, the divergence metric which showed superior results than SKLD and JSD with the stretching function $g(d) = d^{\frac{1}{2}}$, we have investigated the effect of feature selection in the disambiguation procedure; we applied the Shapley Value techniques

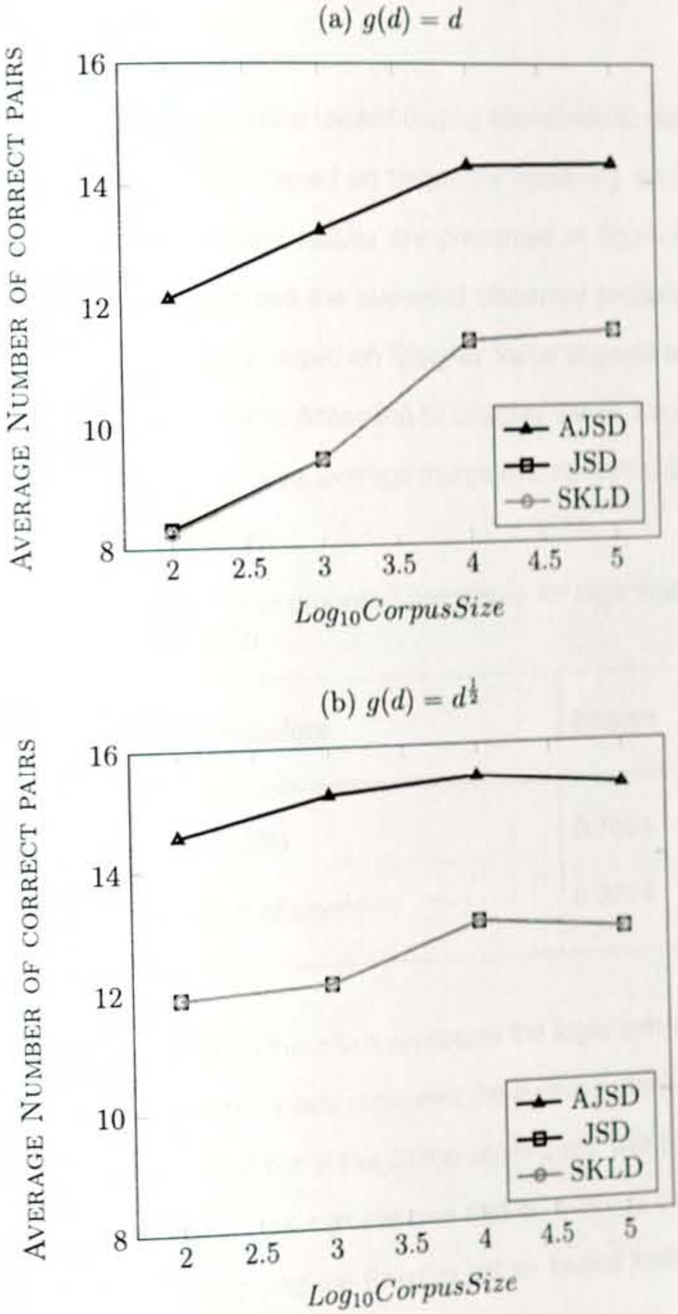

Fig. 6.1 Synthesis of the performance comparison for the three divergence metrics SKL, JSD and AJSD.

on both dataset(synthetic and real world corpora) to identify the most important tags.

The performance of the procedure toward finding cue-words to disambiguate a given tag is compared with FS based on frequently occurring words used by most of the works in literature and results are presented in figure 6.2 and table 6.1. From the table we can see the cue-word discovery procedure in tags disambiguation process using FS based on Shapley Value showed a better performance with F1-score about 93%. According to Shapley Value, the importance of a tag is determined according to its average marginal contribution score which is independent from the frequency.

Table 6.1 F1 Scores-compares the proposed procedure for tags disambiguation with and without feature selection

The Procedure	Scores
With FS based on most FOWs	0.7504
With FS based on Sh value of words	0.9274

From figure 6.2, the labels on the x axis represent the logarithm in base 10 of the size of the corpus and on the y axis represent the average number of correct disambiguation pairs discovered out of the 20 top-score pairs, using the score of equation (4.2) with, $g(d) = d^{\frac{1}{2}}$. One can observe that on average the number of correct pairs discovered by applying the Shapley Valuer based feature selection technique is always higher than the other case.

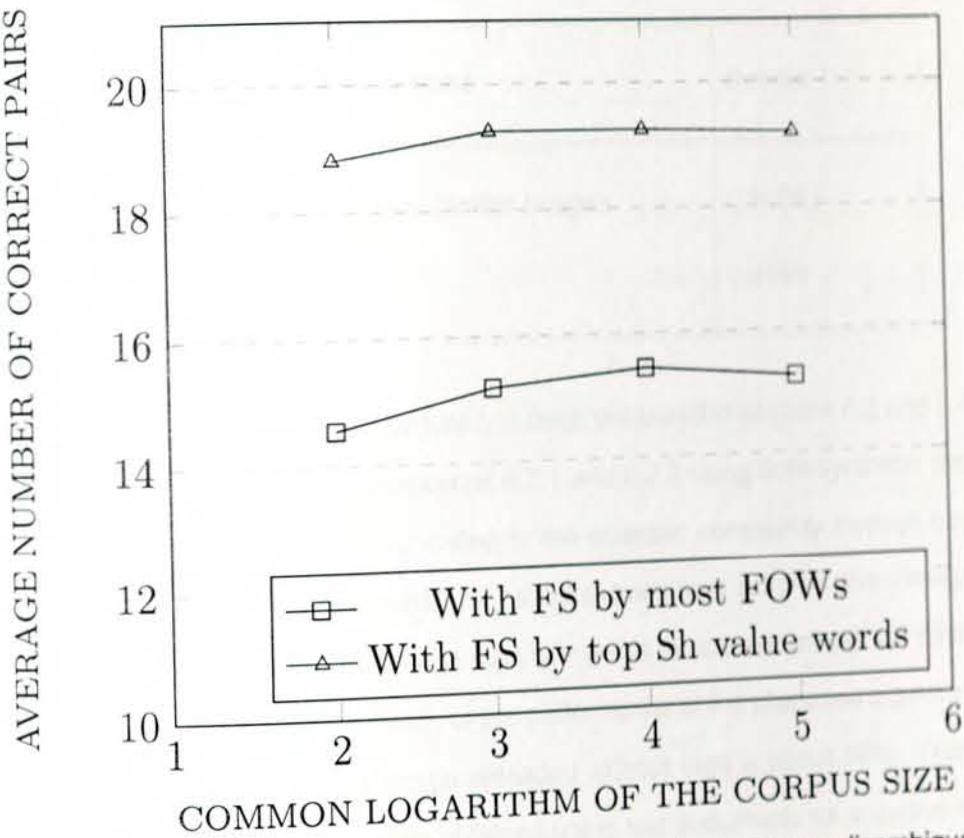

Fig. 6.2 Synthesis of the performance comparison for the two cases disambiguation procedure with feature selection based on the Shapley Value(With FS based on Sh value of words) and with feature selection based on the most frequently occurring words(With FS based on most FOWs).

6.3 Automatic Image Tags generation

The results of the evaluation using F1-score is summarized in table 6.2.

Table 6.2 F1 Score Results to evaluate the performance of the procedure for Automatic Tags Recommendation

Corpus Used	Scores
Tags from Previously Tagged Similar Images	0.9205
Wikipedia Text Document	0.9168

The performance of the procedure has been evaluated in sections 5.3 and 5.4 with the results obtained in subsections 6.2.1 and 6.2.2 using both synthetic and real world datasets and communicated to the scientific community through two articles. Here, the difference is that, we have assumed tags set for a given image as words of a sentence (i.e. image tags are words in one sentence) but other things are the same. The accuracy of the performance of the proposed algorithm to generate tags for the new image uploaded without tags is about 92%. From table 6.2 , we can see the results obtained using text documents as a corpus to determine co-occurring tags is also promising.

The work presented in (Barai and Cardenas, 2010), which is more related to work in regarding to recommending tags of new image automatically, used techniques of finding frequently occurring tags which is up to step 2 of our procedure using previously annotated images and obtained 69% accuracy . Tags relatedness to determine co-occurrence of tags and selecting important tags set in the determining of tags co-occurrence are not performed. Determining tags co-occurrences is important to minimize tag ambiguities.

6.4 Summary of findings

From the above experimental results, the following findings are obtained:

1. In order to determine tags relatedness to determine cue-words of ambiguous tag, a tag-sense disambiguation procedure is proposed. Since, this procedure depends on effective word to context relatedness measures, AJSD is showed as best metric which takes into account statistical fluctuations and at low statistics, i.e. with small corpora.
2. Using the selected divergence measure (i.e. AJSD) together with the Shapley Value feature selection technique about 93% accuracy of correct pairs of tags in the process of tag disambiguation is obtained. The Shapley Value based method to select important feature, is a way of doing feature selection in a form that turns out to be more effective than considering the FS based on FOWs method, i.e. choosing the features based on their performance in isolation. We demonstrated experimentally, the importance of constructing tags probability distributions based on their co-occurrences with the features of the highest Shapley Value scores.
3. Technique of using sentences of texts from semantically rich documents (i.e: Wikipedia) showed in determining tags relatedness.

4. Findings obtained from 1 to 3 are used to recommend tags of an image automatically using tags of previously tagged images.

Chapter 7 Conclusion and Future Works

In this chapter conclusion and future works are presented. First, we summarized main issues of this thesis and then future works and direction resulted from this study are presented.

7.1 Thesis Summary

This thesis investigated the role of tags/words relatedness in tags recommendation for image annotation in folksonomies toward assigning tags to a newly uploaded image: semi-automatically and automatically.

In a typical semi-automatic image tagging, we developed a technique to recommend additional tags to the initial tag supplied by the user who is uploading an image to enrich the semantics of the original tag. These additional tags are recommended to disambiguate and add information in order to minimize tag context scarcity.

If a user doesn't provide initial tag to the image she/he is uploading, this work proposed automatic tags recommendation technique. In this technique, tags of similar images to the target image are analyzed based on tags relatedness information and required tags for the new image are suggested.

In both cases we analyzed word-to-context and word-to-word relatednesses to exploit natural (quasi) co-occurrences of words(words which don't co-occur accidentally). We have used external resource from the web(i.e.; Wikipedia text document) to analyze words co-occurrences. These should rely on effective word-to-context and word-to-word relatednesses metrics. Among the most effective relatedness metrics are those defined on the basis of a feature vector representation of the words.

We used metric derived from a Maximum Likelihood estimator of the Jensen-Shannon Divergence among feature-count histograms-AJSD and we showed that the performance of such a metric in terms of quality of the output is better than both the Jensen-Shannon and the Symmetrized Kullback-Leibler divergence between histograms.

In this thesis, a new algorithm, with corresponding experimental results, for selecting *feature words*, in fact list of words (resulted from pairs of cue-words) that might help a user to disambiguate a freely chosen tag and to find similar tags (or tags of same(similar) image), used to annotate user provided content(tmage) in a web-based social tagging system such as Flickr, is presented. The novelty of the presented algorithm is applying a feature selection approach based on the *Shapley Value*, a metric initially proposed in the domain of Coalitional Game Theory, which measures the "importance" of a component within a coalition. This thesis addresses an interesting open issue in the literature proposing a novel idea in the field of social tagging to determine words/tags relatedness for tag-word disambiguation and tags co-occurrence calculation. Experiments have been performed on both synthetic and real corpus, which show a better accuracy of the results when the Shapley Value metric is used in finding words/tags relatedness.

An experiment, using the synthetic corpus, is conducted by comparing three probability distributions dis(similarity) measures (KLD, JSD and AJSD) to establish the relatedness of one word/tag to another. With respect to the procedure suggested to disambiguate a target tag/word, AJSD always showed better performance. We have used a controlled artificial corpus, in order to make a more precise assessment of the accuracy of the metrics.

The second experiment used a Shapley Value Analysis to calculate the relative weights of tags/words, to select important features set, that was applied to the same corpus for the first experiment to further improve the accuracy of tags disambiguation procedure, while the third experiment applied Wikipedia textual document to implement the procedure on the real-world dataset and results obtained are promising.

The fourth experiment, conducted to evaluate automatic tags generation from previously tagged images on social image sharing web site, Flickr, used the same metrics as that of the second experiment. Wikipedia textual documents are also used to improve the semantics of the results and compared.

Generally, we have answered the research questions raised as follows:

To answer the *first research question*, in a typical semi-automatic tagging technique, a procedure is developed which recommend additional tags for user provided tag . In this procedure when a user contributes an image with tag(s), the tag(s) are checked for ambiguity. If it is found to be ambiguous, then cue-words will be recommended(based on the score obtained from the selected divergence metric) to the user for choice of sense in their mind, automatically. Then the original user provided tag and the new tag will be augmented and assigned as a final annotation of the uploaded image.

To answer the *second research question*, we have developed a procedure which is able to generate tags automatically when the user uploads an image without tags. For unlabeled image, the procedure searches visually similar images from the social media with user contributed tags. The annotations (or tags) are analyzed and ranked based on their relevancies to the given image. Then most relevant set of tags are selected and assigned to the new image.

To answer the *third research question*, a feature selection algorithm is proposed using *Shapley Value Analysis*,a well-known concept in the field of game theory (Hualong and Jing, 2011; Jeffery et al., 2006; Shapley, 1953),used to evaluate the importance of individual feature with respect to the intended task at hand by taking in to considerations the relationships among features. In this procedure a feature is a word and this word will belong to the futures set, if it has highest Shapley Value in the marginal contributions toward the improvement of the accuracy of the procedure.

To answer the *fourth research question*,which will improve the quality of tags, an external textual resources from Wikipedia is used. From English text document dump in Wikipedia a Corpus is extracted using the technique for Text Extractor[1] and used to find words relationship based on words Co-occurrence statistics.

[1]https://blog.afterthedeadline.com/2009/12/04/generating-a-plain-text-corpus-from-wikipedia/, by rsmudge, 2009

7.2 Future Works and directions

This research work tried to establish the approaches required to increase tags accuracy in images. However, the work by no means can be one-time solution and should be extended.

A future work can apply word relatedness metrics and the procedure to select feature words using Shapley Value analysis suggested in this thesis on Wikipedia textual documents (or articles) to focus on sentences which provide relevant and important information about words relationship without doing linguistic annotation process to build folksonomy of words disambiguation, not only a single word but also pair of two words or triple of three words and so on. The technique will collect ambiguous words (possibly as exhaustive as possible) and build folksonomy of these words with their possible senses so that a research work can use this constructed folksonomy in the process of resource organization and retrieval. Besides one can extend this technique for building tag clouds, say TagNet, for words co-occurrences to navigate across resources linked to a search engine.

In social tagging systems (or Folksonomies), we have three important elements: Resources, users, and tags. In this thesis, we have focused on relationships of Tags (particularly tags of images) and partially Images-Tags relations. In the future, this work can be extended to study relationship between Users-Resources, Users-Tags, and Resources-Tags by applying importance of co-occurrences.This will be further extended to recommendation systems by incorporating user-to-user correlation, resource-to-resource correlation, resource-to-user correlation types of recommendation strategies to improve resource organization and retrieval using one of pattern recognition models, for instance Hidden Markov Model.

By analyzing words relatedness information, one can extend this work to extract meta-data on a given domain from folksonomies(Fusion of multiple folksonomies).

Even if the proposed approach is fully statistical, in the future, we would like to explore the effectiveness of our approach for other languages.Different languages have their own syntactic and semantic structure. This work can be extended to words disambiguation technique other than English like 'Amharic word disambiguation.

Furthermore, we would also like to test our method on various domains specific application, for instance, health care recommender systems.

References

Abbasi, R., Chernov, S., Nejdl, W., Paiu, R., and Staab, S. (2009). Exploiting flickr tags and groups for finding landmark photos. In *European Conference on Information Retrieval*, pages 654–661. Springer.

Abbasi, R. and Staab, S. (2008). Introducing triple play for improved resource retrieval in collaborative tagging systems. In *Proceedings of the ECIR Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR'08) March 30, 2008. Barcelona.* Citeseer.

Abel, F., Henze, N., Kawase, R., Krause, D., and Siehndel, P. (2010). Tagme!: Enhancing social tagging with spatial context. In *International Conference on Web Information Systems and Technologies*, pages 114–128. Springer.

Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749.

Afghah, F., Razi, A., and Najarian, K. (2015). A shapley value solution to game theoretic-based feature reduction in false alarm detection. *arXiv preprint arXiv:1512.01680*.

Alelyani, S., Tang, J., and Liu, H. (2013). Feature selection for clustering: A review. *Data Clustering: Algorithms and Applications*, 29:110–121.

Balbi, S. (2010). Beyond the curse of multidimensionality: high dimensional clustering in text mining. *Stat. Appl. J. Appl. Stat.*

Banerjee, M. and Pal, N. R. (2015). Unsupervised feature selection with controlled redundancy (ufescor). *IEEE Transactions on Knowledge and Data Engineering*, 27(12):3390–3403.

Barai, S. and Cardenas, A. F. (2010). Image annotation system using visual and textual features. In *DMS*, volume 10, pages 289–296.

Barclay, L. (2009). Tagging: People-powered metadata for the social web (smith, g.; 2008)[book review]. *IEEE Transactions on Professional Communication*, 52(3):321–322.

Barrios, J. M., Díaz-Espinoza, D., and Bustos, B. (2009). Text-based and content-based image retrieval on flickr. In *Similarity Search and Applications, 2009. SISAP'09. Second International Workshop on*, pages 156–157. IEEE.

Begelman, G., Keller, P., Smadja, F., et al. (2006). Automated tag clustering: Improving search and exploration in the tag space. In *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*, pages 15–33.

Belanche, L. A. and González, F. F. (2011). Review and evaluation of feature selection algorithms in synthetic problems. *arXiv preprint arXiv:1101.2320*.

Benz, D., Grobelnik, M., Hotho, A., Jaschke, R., Mladenic, D., Servedio, V. D., Sizov, S., and Szomszor, M. (2008). Analyzing tag semantics across collaborative tagging systems. In *Dagstuhl Seminar 08391–Working Group Summary*.

Bernardi, R., Cakici, R., Elliott, D., Erdem, A., Erdem, E., Ikizler-Cinbis, N., Keller, F., Muscat, A., and Plank, B. (2016). Automatic description generation from images: A survey of models, datasets, and evaluation measures. *J. Artif. Intell. Res.(JAIR)*, 55:409–442.

Bichler, M. (2006). Design science in information systems research. *Wirtschaftsinformatik*, 48(2):133–135.

Bischoff, K., Firan, C. S., Nejdl, W., and Paiu, R. (2008). Can all tags be used for search? In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 193–202. ACM.

Blei, D. M. and Jordan, M. I. (2003). Modeling annotated data. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 127–134. ACM.

Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and information systems*, 34(3):483–519.

Brin, S. and Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18):3825–3833.

Budanitsky, A. and Hirst, G. (2006). Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32:13–47.

Bulo, S. R., Rabbi, M., and Pelillo, M. (2011). Content-based image retrieval with relevance feedback using random walks. *Pattern Recognition*, 44(9):2109–2122.

Cantador, I., Konstas, I., and Jose, J. M. (2011). Categorising social tags to improve folksonomy-based recommendations. *Web semantics: science, services and agents on the World Wide Web*, 9(1):1–15.

Cattuto, C., Benz, D., Hotho, A., and Stumme, G. (2008). Semantic grounding of tag relatedness in social bookmarking systems. *The Semantic Web-ISWC 2008*, pages 615–631.

Cattuto, C., Loreto, V., and Pietronero, L. (2006). Collaborative tagging and semiotic dynamics. *arXiv preprint cs/0605015*.

Chandrashekar, G. and Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28.

Chang, S.-K. and Hsu, A. (1992). Image information systems: where do we go from here? *IEEE transactions on Knowledge and Data Engineering*, 4(5):431–442.

Chaplot, D. S., Bhattacharyya, P., and Paranjape, A. (2015). Unsupervised word sense disambiguation using markov random field and dependency parser. In *AAAI*, pages 2217–2223.

Chen, J., Zhu, Y.-H., Wang, H.-F., Jin, W., and Yu, Y. (2012). Effective and efficient multi-facet web image annotation. *Journal of Computer Science and Technology*, 27(3):541–553.

Chen, K. (2005). Linear discriminant analysis (lda).

Chen, X., Hu, X., Zhou, Z., Lu, C., Rosen, G., He, T., and Park, E. (2010). A probabilistic topic-connection model for automatic image annotation. In *Proceedings* *of the 19th ACM international conference on Information and knowledge management*, pages 899–908. ACM.

Cohen, S., Ruppin, E., and Dror, G. (2005). Feature selection based on the shapley value. *In other words*, 1:98Eqr.

Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

Curtis, B. A., Nickolas, S. E., and Vicknair, W. E. (2012). Relevant term extraction and classification for wiki content. US Patent 8,196,039.

Datta, R., Li, J., and Wang, J. Z. (2005). Content-based image retrieval: approaches and trends of the new age. In *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*, pages 253–262. ACM.

Dattolo, A., Ferrara, F., and Tasso, C. (2010). The role of tags for recommendation: a survey. In *Human System Interactions (HSI), 2010 3rd Conference on*, pages 548–555. IEEE.

Dellschaft, K. and Staab, S. (2008). An epistemic dynamic model for tagging systems. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 71–80. ACM.

Deshmane, P. and Wankhade, N. (2014). A survey on collaborative tagging.

Dharani, T. and Aroquiaraj, I. L. (2013). A survey on content based image retrieval. In *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on*, pages 485–490. IEEE.

Dobrescu, M., Stoian, M., and Leoveanu, C. (2010). Multi-modal cbir algorithm based on latent semantic indexing. In *Internet and Web Applications and Services (ICIW), 2010 Fifth International Conference on*, pages 37–42. IEEE.

Doerfel, S., Zoller, D., Singer, P., Niebler, T., Hotho, A., and Strohmaier, M. (2016). What users actually do in a social tagging system: a study of user behavior in bibsonomy. *ACM Transactions on the Web (TWEB)*, 10(2):14.

Duda, R. O., Hart, P. E., and Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.

Duygulu, P., Barnard, K., de Freitas, J. F., and Forsyth, D. A. (2002). Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *European conference on computer vision*, pages 97–112. Springer.

Dy, J. G. and Brodley, C. E. (2004). Feature selection for unsupervised learning. *Journal of machine learning research*, 5(Aug):845–889.

Fan, J., Shen, Y., Zhou, N., and Gao, Y. (2010). Harvesting large-scale weakly-tagged image databases from the web.

Feng, S., Manmatha, R., and Lavrenko, V. (2004). Multiple bernoulli relevance models for image and video annotation. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE.

Feng, Y. and Lapata, M. (2008). Automatic image annotation using auxiliary text information. In *ACL*, volume 8, pages 272–280.

Font, F., Serra, J., and Serra, X. (2013). Folksonomy-based tag recommendation for collaborative tagging systems. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 9(2):1–30.

Freund, Y., Iyer, R., Schapire, R. E., and Singer, Y. (2003). An efficient boosting algorithm for combining preferences. *Journal of machine learning research*, 4(Nov):933–969.

Fu, W.-T., Kannampallil, T., Kang, R., and He, J. (2010). Semantic imitation in social tagging. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 17(3):12.

Gemmell, J., Ramezani, M., Schimoler, T., Christiansen, L., and Mobasher, B. (2009). The impact of ambiguity and redundancy on tag recommendation in folksonomies. In *Proceedings of the third ACM conference on Recommender systems*, RecSys '09, pages 45–52, New York, NY, USA. ACM.

Gemmell, J., Shepitsen, A., Mobasher, B., and Burke, R. (2008a). Personalization in folksonomies based on tag clustering. *Intelligent techniques for web personalization & recommender systems*, 12.

Gemmell, J., Shepitsen, A., Mobasher, B., and Burke, R. (2008b). Personalization in folksonomies based on tag clustering. *Intelligent techniques for web personalization & recommender systems*, 12.

Gemmell, J., Shepitsen, A., Mobasher, B., and Burke, R. (2008c). Personalizing navigation in folksonomies using hierarchical tag clustering. In *International Conference on Data Warehousing and Knowledge Discovery*, pages 196–205. Springer.

Gemmell, J., Shepitsen, A., Mobasher, B., and Burke, R. (2008d). Personalizing navigation in folksonomies using hierarchical tag clustering. In *Proceedings of the 10th international conference on Data Warehousing and Knowledge Discovery*, DaWaK '08, pages 196–205, Berlin, Heidelberg. Springer.

Ghodsi, A. (2006). Dimensionality reduction a short tutorial. *Department of Statistics and Actuarial Science, Univ. of Waterloo, Ontario, Canada*, 37:38.

Ghosh, S. and Bandyopadhyay, S. K. (2013). A tutorial review of automatic image tagging technique using text mining. In *International Journal Of Research In Engineering And Technology ISSN*, pages 2319–1163.

Giyanani, R. (2013). A survey on word sense disambiguation. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 14:30–33.

Golub, G. H. and Van Loan, C. F. (2012). *Matrix computations*, volume 3. JHU Press.

Gong, T., Li, S., and Tan, C. L. (2010). A semantic similarity language model to improve automatic image annotation. In *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on*, volume 1, pages 197–203. IEEE.

Goutte, C. and Gaussier, E. (2005). A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *ECIR*, volume 5, pages 345–359. Springer.

Gupta, M., Li, R., Yin, Z., and Han, J. (2010). Survey on social tagging techniques. *ACM Sigkdd Explorations Newsletter*, 12(1):58–72.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182.

Hardoon, D. R., Saunders, C., Szedmak, S., and Shawe-Taylor, J. (2006). A correlation approach for automatic image annotation. In *International Conference on Advanced Data Mining and Applications*, pages 681–692. Springer.

Hardoon, D. R., Szedmak, S., and Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12):2639–2664.

He, X., Cai, D., and Niyogi, P. (2006). Laplacian score for feature selection. In *Advances in neural information processing systems*, pages 507–514.

Hevner, A. and Chatterjee, S. (2010). Design science research in information systems. In *Design research in information systems*, pages 9–22. Springer.

Holland, S. M. (2008). Principal components analysis (pca). *Department of Geology, University of Georgia, Athens, GA*, pages 30602–2501.

Horiuchi, S., Moriguchi, H., Honiden, S., and Shengbo, X. (2013). Automatic image description by using word-level features. In *Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service*, pages 309–314. ACM.

Hotho, A., Jäschke, R., Schmitz, C., and Stumme, G. (2006a). Folkrank: A ranking algorithm for folksonomies. In *LWA*, volume 1, pages 111–114.

Hotho, A., Jäschke, R., Schmitz, C., and Stumme, G. (2006b). Information retrieval in folksonomies: Search and ranking. In *ESWC*, volume 4011, pages 411–426. Springer.

Hualong, B. and Jing, X. (2011). Hybrid feature selection mechanism based high dimensional datasets reduction. *Energy Procedia*, (11):4973–4978.

Iivari, J. (2007). A paradigmatic analysis of information systems as a design science. *Scandinavian journal of information systems*, 19(2):5.

Ivanov, I., Vajda, P., Goldmann, L., Lee, J.-S., and Ebrahimi, T. (2010). Object-based tag propagation for semi-automatic annotation of images. In *Proceedings of the international conference on Multimedia information retrieval*, pages 497–506. ACM.

Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., and Stumme, G. (2007). Tag recommendations in folksonomies. *Knowledge Discovery in Databases: PKDD 2007*, pages 506–514.

Jayaswal, D. and Shrivastava, V. (2015). A literature review on satellite image retrieval techniques. *International Journal of Computer Science Trends and Technology (IJCST)-Volume*, 3.

Jeffery, I. B., Higgins, D. G., and Culhane, A. C. (2006). Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC bioinformatics*, 7:359–359.

Jeon, J., Lavrenko, V., and Manmatha, R. (2003). Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 119–126. ACM.

Jeon, J. and Manmatha, R. (2004). Using maximum entropy for automatic image annotation. In *International Conference on Image and Video Retrieval*, pages 24–32. Springer.

Jesus, R., Abrantes, A. J., and Correia, N. (2011). Methods for automatic and assisted image annotation. *Multimedia Tools and Applications*, 55(1):7–26.

Jiang, J. J. and Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. *arXiv preprint cmp-lg/9709008*.

Jin, R., Chai, J. Y., and Si, L. (2004). Effective automatic image annotation via a coherent language model and active learning. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 892–899. ACM.

Kamoi, Y., Furukawa, Y., Sato, T., Kiwada, Y., and Takagi, T. (2007). Automatic image annotation based on visual cognitive theory. In *Fuzzy Information Processing Society, 2007. NAFIPS'07. Annual Meeting of the North American*, pages 239–244. IEEE.

Ke, X., Li, S., and Cao, D. (2012). A two-level model for automatic image annotation. *Multimedia Tools and Applications*, 61(1):195–212.

Kim, H., Howland, P., and Park, H. (2005). Dimension reduction in text classification with support vector machines. *Journal of Machine Learning Research*, 6(Jan):37–53.

Kim, H.-L., Decker, S., and Breslin, J. G. (2010). Representing and sharing folksonomies with semantics. *Journal of Information Science*, 36(1):57–72.

Kulkarni, A., Gunturu, H., and Datla, S. (2008). Association-based image retrieval. In *Automation Congress, 2008. WAC 2008. World*, pages 1–6. IEEE.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Kumari, B. (2012). Feature subset selection in large dimensionality using correlation based ga-svm. *International Journal of Computer Applications*, 45(6):5–8.

Laaksonen, J., Koskela, M., Laakso, S., and Oja, E. (2000). Picsom–content-based image retrieval with self-organizing maps. *Pattern recognition letters*, 21(13):1199–1207.

Lavrenko, V., Manmatha, R., and Jeon, J. (2003). A model for learning the semantics of pictures. In *NIPS*, volume 1.

Lee, K., Kim, H., Shin, H., and Kim, H.-J. (2009). Tag sense disambiguation for clarifying the vocabulary of social tags. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, pages 729–734. IEEE.

Lee, S., De Neve, W., and Ro, Y. M. (2010). Tag refinement in an image folksonomy using visual similarity and tag co-occurrence statistics. *Signal Processing: Image Communication*, 25(10):761–773.

Leong, C. W., Mihalcea, R., and Hassan, S. (2010). Text mining for automatic image tagging. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 647–655. Association for Computational Linguistics.

Leung, Y., Chang, C., Hung, Y., and Fung, P. (2006). Gene selection for brain cancer classification. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 5846–5849. IEEE.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., and Liu, H. (2016). Feature selection: A data perspective. *arXiv preprint arXiv:1601.07996*.

Li, X., Chen, L., Zhang, L., Lin, F., and Ma, W.-Y. (2006). Image annotation by large-scale content-based image retrieval. In *Proceedings of the 14th ACM international conference on Multimedia*, pages 607–610. ACM.

Li, Y., Geng, B., Zhou, C., and Xu, C. (2011). Learning to combine ad-hoc ranking functions for image retrieval. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 817–822. IEEE.

Lin, D. (1998). Automatic retrieval and clustering of similar words. In *Proceedings of the 17th international conference on Computational linguistics-Volume 2*, pages 768–774. Association for Computational Linguistics.

Lin, J., Yuan, J., Duan, L.-Y., Luo, S., and Gao, W. (2012a). Social image tagging by mining sparse tag patterns from auxiliary data. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, pages 7–12. IEEE.

Lin, Z., Ding, G., Hu, M., Wang, J., and Sun, J. (2012b). Automatic image annotation using tag-related random search over visual neighbors. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1784–1788. ACM.

Lindstaedt, S., Mörzinger, R., Sorschag, R., Pammer, V., and Thallinger, G. (2009). Automatic image annotation using visual content and folksonomies. *Multimedia Tools and Applications*, 42(1):97–113.

Lipczak, M. and Milios, E. (2010). The impact of resource title on tags in collaborative tagging systems. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, pages 179–188. ACM.

Liu, D., Wang, M., Yang, L., Hua, X.-S., and Zhang, H. J. (2009). Tag quality improvement for social images. In *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, pages 350–353. IEEE.

Liu, H. and Motoda, H. (2007). *Computational methods of feature selection*. CRC Press.

Liu, J., Wang, B., Li, M., Li, Z., Ma, W., Lu, H., and Ma, S. (2007a). Dual cross-media relevance model for image annotation. In *Proceedings of the 15th ACM international conference on Multimedia*, pages 605–614. ACM.

Liu, Y., Zhang, D., Lu, G., and Ma, W.-Y. (2007b). A survey of content-based image retrieval with high-level semantics. *Pattern recognition*, 40(1):262–282.

Ljubešić, N., Boras, D., Bakarić, N., and Njavro, J. (2008). Comparing measures of semantic similarity. In *30th International Conference on Information Technology Interfaces, Cavtat.*

Llorente, A., Overell, S., Liu, H., Hu, R., Rae, A., Zhu, J., Song, D., and Rüger, S. (2008). Exploiting term co-occurrence for enhancing automated image annotation. In *Workshop of the Cross-Language Evaluation Forum for European Languages*, pages 632–639. Springer.

Llorente, A., Overell, S., Liu, H., Hu, R., Rae, A., Zhu, J., Song, D., and Rüger, S. (2009). Exploiting term co-occurrence for enhancing automated image annotation. *Lecture Notes in Computer Science*, 5706:632–639.

Llorente, A. and Rüger, S. (2008). Can a probabilistic image annotation system be improved using a co-occurrence approach. In *Workshop on Cross-Media Information Analysis, Extraction and Management at the 3rd International Conference on Semantic and Digital Media Technologies.*

Lowe, D. G. (2013). Distinctive image features from scale-invariant keypoints. 5 de january de 2004. *Cited on*, page 7.

Luo, X. and Zincir-Heywood, A. (2004). Combining word based and word co-occurrence based sequence analysis for text categorization. In *Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on*, volume 3, pages 1580–1585. IEEE.

Lux, M., Pitman, A., and Marques, O. (2010). Callisto: Tag recommendations by image content. *WISMA 2010*, page 87.

Manning, C. and Schütze, H. (1999). *Foundations of statistical natural language processing.* MIT press.

Mante, R., Kshirsagar, M., and Chatur, D. P. (2014). A review of literature on word sense disambiguation. *Int. J. Comput. Sci. Inf. Technol.(IJCSIT)*, 5(2):1475–1477.

March, S. T. and Smith, G. F. (1995). Design and natural science research on information technology. *Decision support systems*, 15(4):251–266.

Martín Wanton, T. and Berlanga Llavori, R. (2012). A clustering-based approach for unsupervised word sense disambiguation.

Mathes, A. (2004a). Folksonomies-cooperative classification and communication through shared metadata.

Mathes, A. (2004b). Folksonomies-cooperative classification and communication through shared metadata. *Computer Mediated Communication*, 47(10).

McParlane, P. J., Moshfeghi, Y., and Jose, J. M. (2014). Collections for automatic image annotation and photo tag recommendation. In *International Conference on Multimedia Modeling*, pages 133–145. Springer.

Metzler, D. and Manmatha, R. (2004). An inference network approach to image retrieval. In *CIVR*, volume 3115, pages 42–50. Springer.

Mika, P. (2005). Ontologies are us: A unified model of social networks and semantics. In *International semantic web conference*, pages 522–536. Springer.

Milicevic, A. K., Nanopoulos, A., and Ivanovic, M. (2010). Social tagging in recommender systems: a survey of the state-of-the-art and possible extensions. *Artificial Intelligence Review*, 33(3):187–209.

Min, H.-s., Choi, J., De Neve, W., Ro, Y. M., and Plataniotis, K. N. (2009). Semantic annotation of personal video content using an image folksonomy. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 257–260. IEEE.

Mori, Y., Takahashi, H., and Oka, R. (1999). Image-to-word transformation based on dividing and vector quantizing images with words. In *First International Workshop on Multimedia Intelligent Storage and Retrieval Management*, pages 1–9. Citeseer.

Mousselly-Sergieh, H., Egyed-Zsigmond, E., Gianini, G., Döller, M., Kosch, H., and Pinon, J.-M. (2013). Tag similarity in folksonomies. In *Proceedings of the XXXI INFORSID congress*, pages 319–334.

Mousselly-Sergieh, H., Egyed-Zsigmond, E., Gianini, G., Döller, M., Pinon, J.-M., and Kosch, H. (2014). Tag relatedness in image folksonomies. *Document numérique*, 17(2):33–54.

Navigli, R. (2009). Word sense disambiguation: A survey. *ACM computing surveys (CSUR)*, 41(2):10.

Niwattanakul, S., Singthongchai, J., Naenudorn, E., and Wanapu, S. (2013). Using of jaccard coefficient for keywords similarity. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 1.

Noll, M. G. and Meinel, C. (2008). Exploring social annotations for web document classification. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 2315–2320. ACM.

Nyberg, E. and Mitamura, T. (2002). Evaluating qa systems on multiple dimensions. In *Proceedings of LREC 2002 Workshop on QA Strategy and Resources*, pages 1–8.

Pal, A. R. and Saha, D. (2015). Word sense disambiguation: A survey. *arXiv preprint arXiv:1508.01346*.

Pal, M. S. et al. (2013). Image retrieval: A literature review. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 1(2):2077–2080.

Pantel, P. and Lin, D. (2002). Discovering word senses from text. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 613–619. ACM.

Papadopoulos, S., Kompatsiaris, Y., and Vakali, A. (2010). A graph-based clustering scheme for identifying related tags in folksonomies. In *Proceedings of the 12th international conference on Data warehousing and knowledge discovery*, DaWaK'10, pages 65–76, Berlin, Heidelberg. Springer-Verlag.

Parveen, A. N., Inbarani, H. H., and Kumar, E. S. (2012). Performance analysis of unsupervised feature selection methods. In *Computing, Communication and Applications (ICCCA), 2012 International Conference on*, pages 1–7. IEEE.

Pasquinelli, M. (2009). Google's pagerank algorithm: A diagram of cognitive capitalism and the rentier of the common intellect. *Deep search*, 3:152–162.

Peddi, B., Xiong, H., ElSherbiny, N., et al. (2010). Evaluation in information retrieval.

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., and Bragge, J. (2006). The design science research process: a model for producing and presenting information systems research. In *Proceedings of the first international conference on design science research in information systems and technology (DESRIST 2006)*, pages 83–106. sn.

Peters, I. and Stock, W. G. (2007). Folksonomy and information retrieval. *Proceedings of the American Society for Information Science and Technology*, 44(1):1–28.

Pino, J. A. (1999). Modern information retrieval. ricardo baeza-yates y berthier ribeiro-neto addison wesley hariow, england, 1999.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3):130–137.

Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation.

Purandare, A. and Pedersen, T. (2004). Word sense discrimination by clustering contexts in vector and similarity spaces. In *CoNLL*, volume 4, pages 41–48.

Qian, M. and Zhai, C. (2013). Robust unsupervised feature selection. In *IJCAI*, pages 1621–1627.

Qian, X., Hua, X.-S., Tang, Y. Y., and Mei, T. (2014). Social image tagging with diverse semantics. *IEEE transactions on cybernetics*, 44(12):2493–2508.

Quintarelli, E. (2005). Folksonomies: power to the people.

Raheja, K. and Gupta, D. (2011). To study the ways to annotate images manual, semi-automatic and fully automatic in m2s and cair. *International Journal of Computer Science and Information Technologies (IJCSIT)*, 2:1725–1728.

Rahuma, A. (2013). Semantically-enhanced image tagging system.

Rajam, F. and Valli, S. (2013). A survey on content based image retrieval. *Life Science Journal*, 10(2):2475–2487.

Robertson, S. (2000). Evaluation in information retrieval. In *Lectures on information retrieval*, pages 81–92. Springer.

Roffo, G. (2016). Feature selection library (matlab toolbox). *arXiv preprint arXiv:1607.01327*.

Ruiz, F. E., Pérez, P. S., and Bonev, B. I. (2009). *Information theory in computer vision and pattern recognition*. Springer Science & Business Media.

Saeys, Y., Inza, I., and Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19):2507–2517.

Sawant, N., Li, J., and Wang, J. Z. (2011). Automatic image semantic interpretation using social action and tagging data. *Multimedia Tools and Applications*, 51(1):213–246.

Seneviratne, L. and Izquierdo, E. (2011). A mathematical approach towards semi-automatic image annotation. In *Signal Processing Conference, 2011 19th European*, pages 559–563. IEEE.

Serrano, R. (2007). Cooperative games: Core and shapley value. encyclopedia of complexity and systems science.

Sevil, S. G., Kucuktunc, O., Duygulu, P., and Can, F. (2010). Automatic tag expansion using visual similarity for photo sharing websites. *Multimedia Tools and Applications*, 49(1):81–99.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317.

Sigurbjörnsson, B. and Van Zwol, R. (2008). Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th international conference on World Wide Web*, pages 327–336. ACM.

Simpson, E. (2008). Clustering Tags in Enterprise and Web Folksonomies. *HP Labs Techincal Reports*.

Singhai, N. and Shandilya, S. K. (2010). A survey on: content based image retrieval systems. *International Journal of Computer Applications*, 4(2):22–26.

Smeulders, A. W., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on pattern analysis and machine intelligence*, 22(12):1349–1380.

Smith, G. (2004). Folksonomy: social classification. *Blog article, August.*

Specia, L. and Motta, E. (2007). Integrating folksonomies with the semantic web. In *Proceedings of the 4th European conference on The Semantic Web: Research and Applications*, ESWC '07, pages 624–639, Berlin, Heidelberg. Springer-Verlag.

Spyrou, E. and Mylonas, P. (2016). A survey on flickr multimedia research challenges. *Engineering Applications of Artificial Intelligence*, 51:71–91.

Su, J.-H., Huang, W.-J., Philip, S. Y., and Tseng, V. S. (2011). Efficient relevance feedback for content-based image retrieval by mining user navigation patterns. *IEEE transactions on knowledge and data engineering*, 23(3):360–372.

Subramanya, S. B. and Liu, H. (2008). Socialtagger-collaborative tagging for blogs in the long tail. In *Proceedings of the 2008 ACM workshop on Search in social media*, pages 19–26. ACM.

Sumathi, T., Devasena, C. L., and Hemalatha, M. (2011). An overview of automated image annotation approaches. *International Journal of Research and Reviews in Information Sciences*, 1(1):1–5.

Tang, J., Alelyani, S., and Liu, H. (2014). Feature selection for classification: A review. *Data Classification: Algorithms and Applications*, page 37.

Thangam, P. S. K. and Angel, R. R. (2013). Semi automatic anno semi automatic annotation exploitation similarity of tation exploitation similarity of pics in personal photo albums n personal photo albums n personal photo albums.

Thangavel, K. and Pethalakshmi, A. (2009). Dimensionality reduction based on rough set theory: A review. *Applied Soft Computing*, 9(1):1–12.

Thielen, P. I. (2010). Social tagging systems–shall we use the collaborative and collec-tive approach to gather competency related information? In *The 3 Rd European Academic Workshop on Electronic Human Resource Management*, pages 186–205.

Tiwari, P. and Kamde, P. (2015). Automatic image annotation and retrieval using contextual information.

Trant, J. (2009). Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, 10(1).

Van Der Maaten, L., Postma, E., and Van den Herik, J. (2009). Dimensionality reduction: a comparative. *J Mach Learn Res*, 10:66–71.

Vander Wal, T. (2007). Folksonomy.

Velmurugan, K. (2014). A survey of content-based image retrieval systems using scale-invariant feature transform (sift). *International Journal of Advanced Research in Computer Science and Software Engineering*, 4.

Virga, P. and Duygulu, P. (2005). Systematic evaluation of machine translation methods for image and video annotation. In *International Conference on Image and Video Retrieval*, pages 174–183. Springer.

Von Alan, R. H., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 28(1):75–105.

Wang, H., Chen, B., and Li, W.-J. (2013). Collaborative topic regression with social regularization for tag recommendation. In *IJCAI*.

Wang, M., Ni, B., Hua, X.-S., and Chua, T.-S. (2012). Assistive tagging: A survey of multimedia tagging with human-computer joint exploration. *ACM Computing Surveys (CSUR)*, 44(4):25.

Wang, S., Tang, J., and Liu, H. (2017). Feature selection.

Weinberger, D. (2005). *Taxonomies to tags: From trees to piles of leaves*. EDventure Holdings.

Weinberger, K. Q., Slaney, M., and Van Zwol, R. (2008). Resolving tag ambiguity. In *Proceedings of the 16th ACM international conference on Multimedia*, MM '08, pages 111–120, New York, NY, USA. ACM.

Wenyin, L., Dumais, S. T., Sun, Y., Zhang, H., Czerwinski, M., and Field, B. A. (2001). Semi-automatic image annotation. In *Interact*, volume 1, pages 326–333.

Wetzker, R., Zimmermann, C., Bauckhage, C., and Albayrak, S. (2010). I tag, you tag: translating tags for advanced user models. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 71–80. ACM.

Willett, P. (2006). The porter stemming algorithm: then and now. *Program*, 40(3):219–223.

Wu, J.-c., Chang, C.-P., and Tsuei, G.-C. (2010). Comparison of feature extraction methods in dimensionality reduction. *Canadian Journal of Remote Sensing*, 36(6):645–649.

Wu, L., Li, M., Li, Z., Ma, W.-Y., and Yu, N. (2007). Visual language modeling for image classification. In *Proceedings of the international workshop on Workshop on multimedia information retrieval*, pages 115–124. ACM.

Wu, L., Yang, L., Yu, N., and Hua, X.-S. (2009). Learning to tag. In *Proceedings of the 18th international conference on World wide web*, pages 361–370. ACM.

Wu, X., Zhang, L., and Yu, Y. (2006). Exploring social annotations for the semantic web. In *Proceedings of the 15th international conference on World Wide Web*, pages 417–426. ACM.

Xu, H., Zhou, X., Wang, M., Xiang, Y., and Shi, B. (2009). Exploring flickr's related tags for semantic annotation of web images. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, page 46. ACM.

Yadav, A. K., Roy, R., Kumar, A. P., et al. (2014). Survey on content based image retrieval and texture analysis with applications. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 7(6):41–50.

Yan, J., Zhang, B., Liu, N., Yan, S., Cheng, Q., Fan, W., Yang, Q., Xi, W., and Chen, Z. (2006). Effective and efficient dimensionality reduction for large-scale and streaming data preprocessing. *IEEE transactions on Knowledge and Data Engineering*, 18(3):320–333.

Yan, R., Natsev, A., and Campbell, M. (2007). An efficient manual image annotation approach based on tagging and browsing. In *Workshop on multimedia* *information retrieval on The many faces of multimedia semantics*, pages 13–20. ACM.

Yang, H.-C., Lee, C.-H., and Chuang, C.-H. (2009). Automatic image annotation using ghsom. In *Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference on*, pages 1188–1191. IEEE.

Yavlinsky, A., Schofield, E., and Rüger, S. (2005). Automated image annotation using global features and robust nonparametric density estimation. In *International Conference on Image and Video Retrieval*, pages 507–517. Springer.

Yin, D., Xue, Z., Hong, L., and Davison, B. D. (2010). A probabilistic model for personalized tag prediction. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 959–968. ACM.

Yu, L. and Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pages 856–863.

Zhang, D., Islam, M. M., and Lu, G. (2012). A review on automatic image annotation techniques. *Pattern Recognition*, 45(1):346–362.

Zhang, X., Li, Z., Wang, S., Yang, Y., and Lv, X. (2015). Location prediction of social images via generative model. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pages 275–282. ACM.

Zhao, Z. and Liu, H. (2007). Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the 24th international conference on Machine learning*, pages 1151–1157. ACM.

Zhou, N., Cheung, W. K., Qiu, G., and Xue, X. (2011). A hybrid probabilistic model for unified collaborative and content-based image tagging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1281–1294.

Zhou, N., Cheung, W. K., Xue, X., and Qiu, G. (2008). Collaborative and content-based image labeling. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE.