International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
1st International Conference on Advent Trends in Engineering, Science and Technology
"ICATEST 2015", 08 March 2015

Web Mining Image Retrieval using Multimodal Fusion Method

Rima P. Lingawar , Milind V. Srode ,Mangesh M. Ghonge
Dept of Computer Science & Engineering
M.E. Student of college of Jagdhambha Engineering & Technology, Yavatmal
Computer Science HOD of Jagdhambha college of Engineering & Technology, Yavatmal
Assistant Professor Jagdambha college of Engineering & Technology, Yavatmal
rimalingawar@yahoo.com

Abstract: The retrieving method utilizes the fusion of the images' multimodal information (textual and visual) which is a recent trend in image retrieval researches. It combines two different data mining techniques to retrieve semantically related images: clustering and association rules mining algorithm. The semantic association rules mining is constructed at the offline phase where the association rules are discovered between the text semantic clusters and the visual clusters of the images to use it later at the online phase. The proposed method will achieve the best precision among different query categories.

Keywords— Image retrieval, Multimodal information fusion, Association rules mining, Clustering.

1. INTRODUCTION

Today, the World Wide Web is the popular and interactive medium to disseminate information. The Web is huge, diverse and dynamic, it contains vast amount of information and provides an access to it at any place at any time. The most of the people use the internet for retrieving information. But most of the time, they gets lots of insignificant and irrelevant document even after navigating several links.

For retrieving information from the Web, Web mining techniques are used. As the demand for image retrieval and browsing online is growing dramatically, there are hundreds of millions of images available on the current World Wide Web. For multimedia documents, the typical keyword-based retrieval methods assume that the user has an exact goal in mind in searching a set of images whereas users normally do not know what they want, or the user faces a repository of images whose domain is less known and content is semantically complicated. In these cases it is difficult to decide what keywords to use for the query. Based on the visual and textual-based image retrieval models, the response to the image query is very much easy and effective. Image retrieval could relay purely on textual metadata which produce lot of garbage in the results because users usually enter that metadata manually which is inefficient, expensive and may not capture every keyword that describes the image. On the other hand, the Content Based Image Retrieval (CBIR) systems can filter images based on their visual contents such as colors, shapes, textures or any other information that can be derived from the image itself which may provide better indexing and return more accurate results [2]. At the same time, these visual features contents extracted by the computer may be different from the image contents that people understand. The fundamental difference between content-based and text-based retrieval systems is that the human interaction is an indispensable part of the latter system. Humans tend to use high- level features (concepts), such as keywords, text descriptors, to interpret images and measure their similarity. While the features automatically extracted using computer vision techniques are mostly low-level features (color, texture, shape, spatial layout, etc.).

The given method is a Multimodal Fusion method based on Association Rules mining (MFAR). It is considered as a late fusion. This method combines two different data mining techniques for retrieving: clustering and association rules mining (ARM) algorithm. It uses ARM algorithm to explore the relations between text semantic clusters and image visual features clusters. The method consists of two main phases: offline and online phase. In the offline phase, the relations among the clusters will be

identified from different modalities to construct the semantic Association Rules (ARs). On the other hand, the online phase (retrieving phase) uses the generated ARM, to retrieve the related images of the query.

2. RELATED WORK

A recent trend for image search is to fuse the two basic modalities of Web images, i.e., textual features (usually represented by keywords) and visual features for retrieval. The fusion of the image visual and textual features was performed in different levels of the retrieving process in image retrieval systems which are early fusion, late fusion, trans-media fusion and re-ranking.

A. Early Fusion

The early fusion approach consists in representing the multimedia objects in a multimodal feature space designed via a joint model that attempts to map image based features with text based features. The simplest early fusion method consists in concatenating both image and text feature representations, after extracting the low level visual features, the extracted vectors are concatenated into one vector to form one unique feature space. The advantage of this strategy is that it enables a true multimedia representation for all the fused modalities where one decision rule is applied on all information sources. Early fusion could be used without feature weighting such in [5]; they concatenate the normalized feature spaces of the visual and the textual features. The main drawback of early fusion is the dimensionality of the resulting feature space which is equal to the sum of all the fused subspaces and that leads to the well-known problem the "curse of dimensionality". Also, increasing in the number of modalities and the high dimensionality make them difficult to learn the cross correlation among the heterogeneous features.

B. Late Fusion

Late fusion strategies do not act at the level of one representation for all the media features but rather at the level of the similarities among each media. In the late fusion, the extracted features of each modality are classified using the appropriate classifier then each classifier provides the decision. Unlike early fusion, where the features of each modality may have different representation, the decisions usually have the same representation. As a result, the fusion of the decisions becomes easier. In addition, it allows for each modality to use the most suitable methods for analyzing and classifying which provides much more flexibility than the early fusion. The main disadvantage of this strategy is that it fails to utilize the feature level correlation among modalities. Also, using different classifiers and different learning process is expensive in term of time and learning for each modality. Late fusion was used widely in image retrieval systems and there is a diversity in the proposed methods.

C. Trans-media Fusion

In this method, the main idea is to use first one of the modalities (say image) to gather relevant documents (nearest neighbors from a visual point of view) and then to use the dual modalities (text representations of the visually nearest neighbors) to perform the final retrieval. Most proposed methods under this category are based on adopted relevance feedback or pseudo-relevance feedback technique .

D. Image Re-ranking

Another level for fusing the visual and textual modalities called as image re-ranking. It consists of two phases: first a text search is used, then the returned list of images is reordered according to the visual feature similarity. While [7] method deals with the clusters of the modalities, [8] proposed a method that construct a semantic relation between text (words) and visual clusters using the ARM algorithm.

3. ANALYSIS OF PROBLEM

Currently, most Web based images search engines rely purely on textual metadata. That produces a lot of garbage in the results because users usually enter that metadata manually which is inefficient, expensive and may not capture every keyword that describes the image. On the other hand, the Content Based Image Retrieval (CBIR) systems can filter images based on their visual contents such as colors, shapes, textures or any other information that can be derived from the image itself which may provide better indexing and return more accurate results. At the same time, these visual features contents extracted by the computer may be different from the image contents that people understand. It requires the translation of high-level user perceptions

into low-level image features and this is the so-called "semantic gap" problem. This problem is the reason behind why the CBIR systems are not widely used for retrieving Web images. A lot of efforts have been made to bridge this gap by using different techniques. In [1], the authors identified the major categories of the state-of the- art techniques in narrowing down the semantic gap one of them is fusing the retrieval results of multimodal features. Fusion for image retrieval (IR) is considered as a novel area with very little achievements in the early days of research [2].In Web medium, the representation of images can be naturally split into two or more independent modalities such as visual features (color, shape…etc) and textual features (metadata and related text, there is need to narrow the semantic gap problem and enhance the retrieval performance by fusing the two basic modalities of Web images, i.e. textual and visual features for retrieving.

4. BASIC OF ASSOCIATION RULES MINING ALGORITHM

ARM is a data mining technique useful for discovering interesting relationships hidden in large databases. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories [6]. The classical example is the rules extracted from the content of the market baskets. Items are things we can buy in a market, and transactions are market baskets containing several items. The collection of all transactions called the transactions database.

There are two important basic measures for association rules, support(s) and confidence(c). Since the database is large and users concern about only those frequently purchased items, usually thresholds of support and confidence are predefined by users to drop those rules that are not so interesting or useful. The two thresholds are called minimal support and minimal confidence respectively, additional constraints of interesting rules also can be specified by the users.

Support(s) of an association rule is defined as the percentage/fraction of records that contain X ∪ Y to the total number of records in the database. The count for each item is increased by one every time the item is encountered in different transaction T in database D during the scanning process. It means the support count does not take the quantity of the item into account. For example in a transaction a customer buys three bottles of water but we only increase the support count number by one.

Confidence of an association rule is defined as the percentage/fraction of the number of transactions that contain X ∪ Y to the total number of records that contain X, where if the percentage exceeds the threshold of confidence an interesting association rule X $\longrightarrow$Y can be generated. Confidence is a measure of strength of the association rules, suppose the confidence of the association rule X$\longrightarrow$*Y* is 80%, it means that 80% of the transactions that contain X also contain Y together, similarly to ensure the interestingness of the rules specified minimum confidence is also pre-defined by users.

The problem of mining association rules is stated as following: *I*= {*i* , *i2* , ... , *im*} is a set of items, *T*={*t1* , *t2* , ... , *tn*} is a transaction database or a set of transactions, each of which contains items of the itemset *I*. Association rule mining is to find out association rules that satisfy the pre- defined minimum support and confidence from a given database [Agrawal and Srikant 1994]. The problem is usually decomposed into two sub problems. One is to find those itemsets whose occurrences exceed a predefined threshold in the database, those itemsets are called frequent or large itemsets. The second problem is to generate association rules from those large itemsets with the constraints of minimal confidence.

In the association rule of the form *X* $\longrightarrow$*Y*, *X* would be called the antecedent or the left hand side, *Y* the consequent or also called the right hand side, as well. It is obvious that the value of the antecedent implies the value of the consequent. The process of mining association rules consists of two main steps. The first step is to identify all the itemsets contained in the data that are adequate for mining association rules. To determine that the itemset is frequent, it should satisfy at least the predefined minimum support count. To measure the support for an itemset, the following formal definition is used:

$$\text{Support}(X) = \text{count}(X) \,/\, (N)$$

Where, *N* is the total number of transactions in the transaction database *T* i.e. *N* = count(*T)*. The second step is to generate rules out of the discovered frequent itemsets. For doing so, a minimum confidence has to be defined. The formal definition to

calculate the rule confidence is given by the following equation:

$$Conf(X \longrightarrow Y) = Count(X \cup Y) / count(X)$$

The confidence of the rule $X \longrightarrow Y$ is a measurement that determines how frequently items in Y appear in transactions that contain X. Different algorithms attempt to allow efficient discovery of frequent patterns and for strong ARs such as the famous Apriori algorithm which will be used later in the proposed method. Apriori is a great improvement in the history of asso-ciation rule mining, There are two processes to find out all the large itemsets from the database in Apriori algorithm. First the candidate itemsets are generated, then the database is scanned to check the actual support count of the corresponding itemsets.

5. PROPOSED WORK

MFAR method consists of two main phases: online phase and offline phase. The next sub sections describe in details the inputs, the outputs and the steps of each phase.

A. Offline Phase

The inputs of this phase are the images dataset which contains two modalities: the images and their associated text. First, the visual and the textual features are extracted to run the clustering algorithm independently over them. Then, the modified association algorithm will identify the relations among the clusters from each modality to construct the ARs (see fig. 1).

- **Features extraction**

The features are selected to balance the color and the edge properties of the images. After extracting the visual features, the images of the dataset are represented as objects in multidimensional space models separately for each visual feature. For textual features, we need to perform several linguistic preprocessing steps (tokenization, removing stop word). Then, each document is described by a vector of the constituent terms –which are the extracted features– that represents the frequency occurrence of each term in the document. The set of all vectors construct the vectorspace model, usually known as a bag-of-words model.

- **Clustering**

For the visual features, the large quantity of images and the high dimensionality of descriptors need for an efficient clustering (or indexing) method. The high dimensional index technique called NOHIS (Non Overlapping Hierarchical Index Structure) is used in the system. After running NOHIS [9] algorithm, the NOHIS-tree is constructed; it is a not balanced binary tree. Then, an adapted k-nearest neighbors search is used for retrieving.

Association rules mining algorithm of clusters

In our case, the items set is the generated images clusters based on the text and based on the visual features l. After quantifying the features space of each modal, we aim to associate the text clusters and the visual feature clusters. Thus, we need to construct the transaction database *T* first to run the ARM algorithm over it.

Frequent Itemsets Mining Algorithm Based On Apriori

Input:

a) The transaction database *T*

b) minsup threshold

Output:

The list of frequently itemsets L

Method

- Find all frequent 2-itemsets with common textual and visual feature
- It uses level wise search where k itemset used to generate (k+1)
- Count support for each candidate by scanning the database
- Eliminate the candidates that are infrequent
- The item set that is not frequent is not subset of k itemset and used to determine AR.

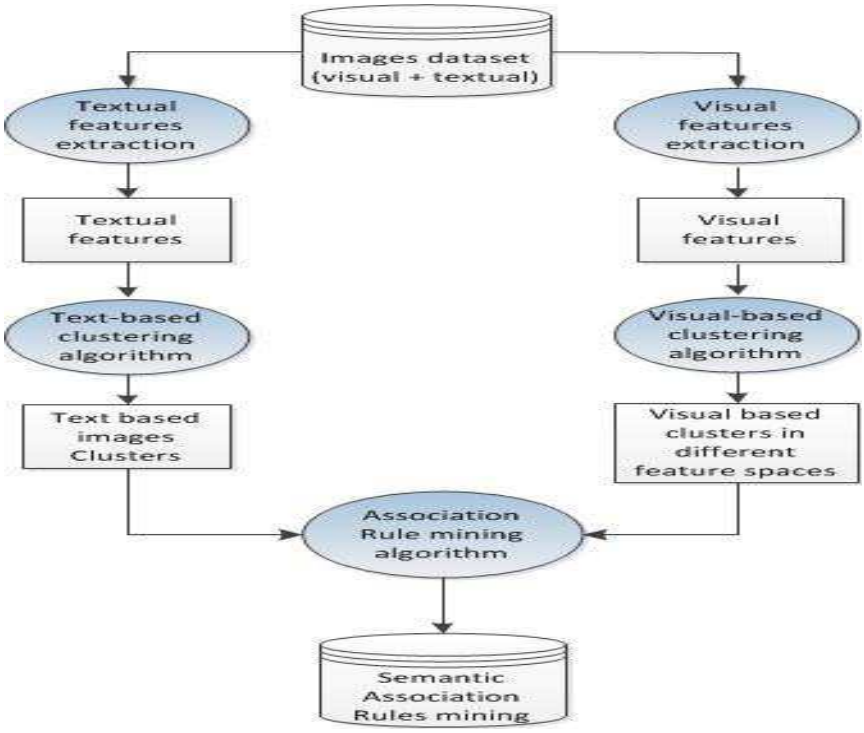

Fig. 1.The offline phase

B. Online phase

It is the retrieving phase. The main retrieving processes are illustrated in fig. 2. The next sections describe each process in this phase.

- **Query Modalities and Processing**

The used query paradigm in this method is the composite paradigm. The basic query model used here is query by example image since when image is used as query, all the information it contains is provided to the system. Using a keyword as query is optional. It could be provided to the system to support the results that generated by the image query. For query image, we need to extract the same visual features that have been extracted from the images of the dataset. For the optional keyword query, we used one keyword and simple text matching to simplify this step.

- **Search and retrieve the related visual clusters**

In addition to the resulted strong ARs from the offline phase, there is another output from that phase: NOHIS-tree which was discussed previously. We need to use the same index trees to retrieve the relevant clusters to the query image for each visual descriptor. In our case, we have two different NOIHS-trees for two different feature spaces. For each feature, the query vector *q* will be used to search in the trees and to retrieve the relevant clusters of *q*. Relevant cluster is a leaf cluster that contains nearest neighbor(s) objects to *q*. The used similarity measurement of images is simply defined as the Euclidean distance between two vectors.

- **Retrieve ARs with similar visual clusters**

It gets the list of the related visual clusters as input and then to make a search in the ARM to get the rules that contain these clusters. If the keyword query was provided, the retrieved rules should be filtered to pick the rules which contain text clusters that have similar term to the text query using simple text match. Then, the images' scores in those text clusters should be increased. The dashed arrow in fig. 2 indicates that it is an optional path.

- **Get images of the text clusters, normalize scores, fuse them and reorder the list**

For all the retrieved ARs, we need to get the images of the text-based clusters. For each image, the relevant score to the query image *q* should be calculated. Since the relevance scores generated from different feature spaces, it is important to normalize the scores before fusing them. In our case, we will use Zero-One linear method which maps the scores into the range of [0, 1] [11]. The normalized scores of different modalities should be fused using CombSum method [3]. Then, if there is a keyword query as input, the fused score of each image that correlated to term matched the query should be increased. Finally, the final fused list will be reordered based on the fused scores.

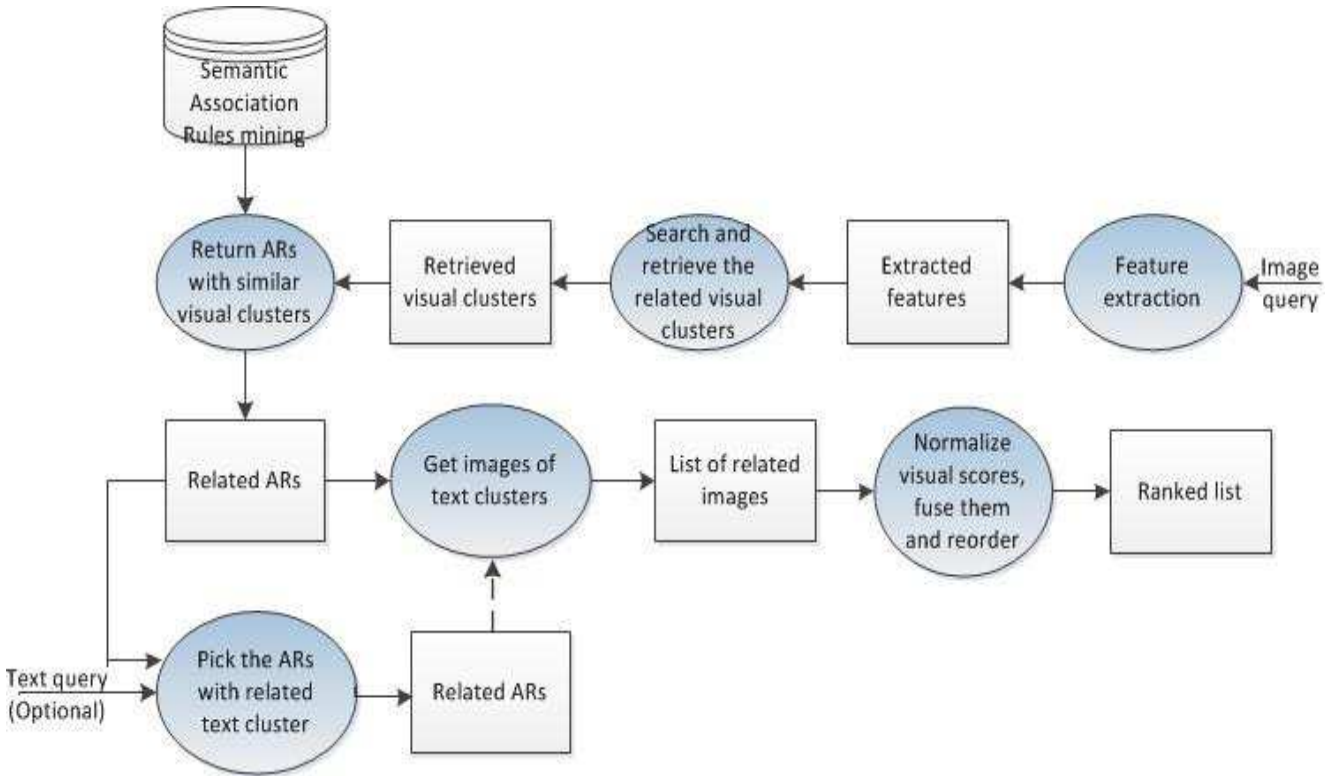

Fig. 2. The online phase

6. CONCLUSION

In this seminar, association rules mining algorithm in our Web image retrieval system to construct a semantic relations between images clusters based on the visual features and the textual features for the same dataset. After constructing the ARs in the offline phase, the retrieving process should starts with example image query in the online phase. The method gives the ability to retrieve images that are semantically related by using the extracted visual features of the query image and by exploring the related ARs from the mining

REFERENCES

[1] Y. Liu, D. Zhanga, G. Lua, and W-Y. Ma , "A survey of content-based image retrieval with high-level semantics", Pattern Recognition, Vol. 40, No. 1. (2007), pp. 262-282.

[2] R. Datta, D. Joshi, J. LI, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age", ACM Computing Surveys (CSUR), April 2008, 40(2):1-60.

[3] S. Wu and S. McClean, "Performance prediction of data fusion for information retrieval".

Information Processing, Management, 2006. 42(4): p. 899-915.

[4] H. Müller, P. Clough, Th. Deselaers, B. Caputo, ImageCLEF (ser. The Springer International Series on Information Retrieval), vol. 32, pp.95 - 114, 2010, Springer-Verlag.

[5] T. Deselaers, T. Weyand, and H. Ney, "Image retrieval and annotation using maximum entropy", in Evaluation of Multilingual and Multi modal Information Retrieval, 2007, pp. 725-734.

[6] Qiankun Zhao Nanyang Technological University, Singapore and Sourav S. Bhowmick Nanyang Technological University, Singapore" Association Rule Mining: A Survey"

[7] S. Wei , Y. Zhao , Z. Zhu , N. Liu, "Multimodal Fusion for Video Search Reranking", IEEE Transactions on Knowledge and Data Engineering, 2010, v.22 n.8, p.1191-1199.

[8] R. He, N. Xiong, L. Yang, J. Park, "Using multi-modal semantic association rules to fuse keywords and visual features automatically for web image retrieval". In: International conference on information fusion. 2011

[9] M. Taileb, S. Lamrous and S. Touati, "Non Overlapping Hierarchical Index Struture", International Journal of Computer Science, vol. 3 no. 1, pp. 29-35, 2008.

[10] T. Tsikrika, A. Popescu, J. Kludas, "Overview of the Wikipedia Image Retrieval Task at ImageCLEF 2011". In: Working Notes of CLEF 2011, Amsterdam, The Netherlands. 2011

[11] S. Wu, "Data Fusion in Information Retrieval". 2012, Springer, Heidelberg.