ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

A Personalized Profile Interface on Web Image Search Using Feedback Mechanism

R.Tamilselvi[1]

Research Scholar, Department of Computer Science, Dr.SNS Rajalakshmi college of Arts and Science, Coimbatore, Tamil Nadu, India [1]

ABSTRACT: Web personalization is the process of transforming a Web site to the needs of particular users, taking advantage of the knowledge obtained from the analysis of the user's navigational behavior in correlation with other information collected in the Web context (structure, content and user profile data). Principal elements of Web personalization include (a) the categorization of Web data, (b) the extraction of correlations between and across different kinds of such data, and (c) the determination of the actions recommended by personalization system. Along with the visual semantic space, a new scheme can be proposed with user preferences in profiles where images can be given rating as positive or negative. During next retrieval or search with similar keywords or images, the negative result images will not be displayed based on rating. This can be done with personalized user profiles for efficient data retrieval of images extraction from the web. Personalized search is an emerging technique to improve the search result relevance at every user level by storing the user search data at web logs as private to that user. This approach not only suggests the personal relevance, but also concentrates on public relevance. To this end, first create local repository for bag-of-words from a small neighborhood around each key point to exploit local structural information which user needs. Adapt the Search Result Clustering Algorithm to discover the visual patterns and distinguish them based on user interests.

KEYWORDS: CBIR, Images, Local User Profiles, Feedback mechanism

I. INTRODUCTION

The revolutionary internet and digital technologies have imposed a need to have a system to organize abundantly available digital images for easy categorization and retrieval. The need to have versatile and general purpose image retrieval (IR) system for a very large image database has attracted focus of many researchers of information technology-giants and leading academic institutions for development of IR techniques .These techniques encompass diversified areas, viz. image segmentation, image feature extraction, representation, mapping of features to semantics, storage and indexing, image similarity-distance measurement and retrieval - making IR system development a challenging task. One of the main problems highlighted was the difficulty of locating a desired image in a large and varied collection. While it is perfectly possible to identify a desired image from a small collection simply by browsing, more effective techniques are needed with collections containing thousands of items. The goal of this proposed framework is to survey the core concepts and techniques in the large subset of cluster analysis with its roots in statistics and decision theory.

A. CONTINUOUS INNOVATION

Content-Based Image Retrieval (CBIR)

CBIR is a technology that in principle helps organize digital image archives according to their visual content. This system distinguishes the different regions present in an image based on their similarity in color, pattern, texture, shape, etc. and decides the similarity between two images by reckoning the closeness of these different regions.

Visual Information Retrieval (Images)

Visual info Retrieval (VIR) may be a comparatively new field of analysis in engineering and Engineering. As in standard info retrieval, the aim of a VIR system is to retrieve all pictures (or image sequences) that are relevant to a user question whereas retrieving as few non-relevant images as attainable. The interpretation method involves extracting (semantic) info from the documents (images) and victimization this info to match the user desires. Image

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

mining deals with the extraction of implicit knowledge, that is, image data relationship or other patterns not explicitly stored in the images. Relevance feedback (RF) given by users refers to a set of approaches learning from a range of users' browsing behaviors on image retrieval.

B. CLUSTERING

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

1.2.1 Types of Clustering Methods

Clustering methods don't require pre-defined categories as in classification methods. Thus, they are more adaptive for various queries.

Hierarchical clustering proceeds successively by either merging smaller clusters into larger ones, or by splitting larger clusters. Hierarchical methods construct the clusters by recursively partitioning the instances in either a top-down or bottom-up fashion.

Partition clustering, on the other hand, attempts to directly decompose the data set into a set of disjoint clusters. The criterion function that the clustering algorithm tries to minimize may emphasize the local structure of the data, as by assigning clusters to peaks in the probability density function, or the global structure. Partitioning methods relocate instances by moving them from one cluster to another, starting from an initial partitioning.

1.2.3 Use of Clustering in Data mining

- **Scalability -** We need highly scalable clustering algorithms to deal with large databases.
- **Ability to deal with different kind of attributes -** Algorithms should be capable to be applied on any kind of data such as interval based (numerical) data, categorical, binary data.
- **Discovery of clusters with attribute shape -** The clustering algorithm should be capable of detect cluster of arbitrary shape. It should not be bounded to only distance measures that tend to find spherical cluster of small size.
- **High dimensionality -** The clustering algorithm should not only be able to handle low- dimensional data but also the high dimensional space.
- **Ability to deal with noisy data -** Databases contain noisy, missing or erroneous data. Some algorithms are sensitive to such data and may lead to poor quality clusters.
- **Interpretability -** The clustering results should be interpretable, comprehensible and usable.

II. LITERATURE REVIEW

Recently a dozen of image/video re ranking methods has been proposed to exploit the usage of the visual information for refining the text-based search result. Most of these re ranking methods utilize the visual information in an unsupervised and passive manner. Unsupervised re ranking methods can only achieve limited performance improvements. This is because the visual information is insufficient to infer the user's intention, especially when the query term is ambiguous.

The problem of clustering search results has been investigated in a number of previous works. Some of them (e.g. [3][6][11][10]) apply traditional clustering algorithms which first cluster documents into topically-coherent groups according to content similarity, and generate descriptive summaries for clusters. However, these summaries are often unreadable, which make it difficult for Web users to identify relevant clusters.

Zamir and Etzioni [16][17] presented a Suffix Tree Clustering (STC) which first identifies sets of documents that share common phrases, and then create clusters according to these phrases. Our candidate phrase extraction process is similar to STC but we further calculate several important properties to identify salient phrases, and utilize learning methods to rank these salient phrases.

Du et al. [12] propose a new study, in which web pages for search engine results are classified as low-adjacence or high-adjacence sets. To match user queries with web pages using formal concept analysis, a concept lattice of the low-adjacence set is defined and the non-redundancy association rules defined by Zaki for the concept lattice are extended. To avoid returning irrelevant web pages for search engine results, technologies that match user queries to web pages have been widely developed.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Nebot et al. [13] present a novel method for mining association rules from semantic instance data repositories expressed in RDF/S and OWL. They take advantage of the schema-level knowledge encoded in the ontology to derive just the appropriate transactions which will later feed traditional association rules algorithms. This process is guided by the analyst requirements, expressed in the form of a query Pattern experiments performed on real world semantic data enjoy Promising results and slimy the usefulness.

Cui et al. [5, 4] classified the query images into eight predefined intention categories and gave different feature weighting schemes to different types of query images. However, it was difficult for only eight weighting schemes to cover the large diversity of all the web images. It was also likely for a query image to be classified to a wrong category.

Kumar et al. have built an image search engine [15] where users can retrieve images of faces based on queries involving multiple visual attributes. However, these methods do not consider the fact that attributes are highly correlated. For example, a person who has a mustache is almost definitely a male, or a person who is Asian is unlikely to have blonde hair.

Rasiwasia et al. [14] mapped visual features to a universal concept dictionary for image retrieval. Farhadi et al. [1] learned part and attribute detectors which were shared across categories and modeled the correlation. Kuo et al. [2] recently augmented each image with relevant semantic features through propagation over a visual graph and a textual graph which were correlated. Parikh and Grauman [7] proposed relative attributes to indicate the strength of an attribute in an image with respect to other images.

Lavrenko et al [8] proposed a continuous relevance model, which directly associates continuous features with words and achieved significant improvement in performance. Jeon et al [9] extended this model and built it with 56,000 Yahoo! news images with noisy annotations and a vocabulary of 4,073 words. This is the largest vocabulary ever proposed, and they discussed noisy annotation filtering and speeding-up schemes.

As a different kind of approach, Pan et al [19] leveraged a graph structure. They constructed a two-layer graph whose nodes are images and their associated captions, and proposed a random-walk-with-restart algorithm to estimate the correlations between new images and the existed captions. Then, in another work of them [20], they extended the model to a three-layer graph with image regions added. In contrast, a few researchers proposed discriminative models. Chang et al [24] learned an ensemble of binary classifiers, each for a specific label. Li et al [18] proposed a Confidence-based Dynamic Ensemble model which is a two-stage classifier.

Carneiro et al [22] attempted to establish a one-to-one mapping between semantic classes and sets of images with the criterion to minimize the error rate. Mori et al [21] uniformly divided each image into sub-images with key words, and then applied vector quantization onto visual features of the sub-images to estimate which words will be assigned to the new image. All the work mentioned above require a supervised learning stage. Hence, the generalization capability is a crucial assessment to their effectiveness.

III. RESEARCH METHODOLOGY

A.WEB PERSONALIZATION TECHNIQUE FOR LOCAL USER PROFILES

The personalization can be enhanced by categorizing the results according to the types. Thus after building the knowledge base, the system can give use recommendation based on the similarity of the user interested domain and the user query. The recommendation procedure of the System has two steps:

The system gives user a list of interested domains. Detect user's current interested domain.

Based on user's current interested domain and combined his or her profile, the system will give him or her set of URLs with ranking scores.

In this way, the system could help the user to retrieve his or her potential interested domains. Besides, a user can change his or her current interested domain by clicking the interested domain list on the same page but with more convenience. In the beginning, if the user does not have a profile in the database, the system displays the user available domains, and then keeps a track of the user's selections. The main advantages of this proposed system are

Collecting information from users to obtain the specified semantic space.

Use relevance feedback mechanism from users and personalize their web image search in local user profile.

Localizing the visual characteristics of the user's intention in this specific semantic space.

The proposed method helps to improve the results of image search engines on the Internet to satisfy the users who desire to see the relevant images in the first few pages. The results of the text based systems that use only the accompanied text of the images, are re-ranked by considering the relevance feedback given by the user earlier by incorporating the visual similarity of the resulting images. It observe that, in general, together with many unrelated

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

ones, the result of text based systems include a subset of correct images, and this set is the largest most similar one compared to other possible subsets.

B. PROCESS OF WEB PERSONALIZED SEARCH USING FEEDBACK MECHANISM

Human perception of image similarity is subjective, semantic, and task-dependent. Although content-based methods hold promises for image retrieval generally, retrieval results based on similarities of pure visual features may not always be meaningful in a perceptual or semantic way. Also, each type of visual feature tends to capture only one aspect of image property and it is usually hard for a user to specify clearly how different aspects are combined. With relevance feedback, it is possible to establish a link between high level concepts and low-level features. Relevance feedback is a supervised active learning technique used to improve the effectiveness of information systems. The main idea is to use positive and negative examples from the user to improve system performance. For a given query, the system first retrieves a list of ranked images according to a predefined similarity metrics. The user marks the retrieved images as relevant (positive examples) to the query or not relevant (negative examples). The system will refine the retrieval results based on the feedback and present a new list of images to the user. Thus the key issue in relevance feedback is how to incorporate positive and negative examples to refine the query and/or to adjust the similarity measure.

User Profiling

The formal representation of user preferences P :

$$P = \{U^+, U^-\}$$

----→ (Eq.1)

whereU^+ ,U^- refer to the set of positive and negative preferences, respectively.
Following the sum notation for sets U^+ and U^- are defined as follows:

$$U^+ = \{U_j^+\},\ j \in N_k,\ U^- = \sum c_i / p_i^-,\ i \in N_n,\ n = |C|$$

----→ (Eq.2)

k is the count of distinct positive preferences contained in the user profile,
C is the set of defined concepts in user profile,
p_{ij}^+ is the degree of participation of concept $c_i$in U_j^+,
$p_i^-$is the degree of participation of concept $c_i$in U^- and
$U_j^+ = \sum c_i / p_{ij}^+$, i € N_n, j € N_k, n =| C |

This definition allows participation of a single concept in multiple preferences and to different degrees. As all relations existing in the local repository are defined on the set C of concepts, we define user preferences on the same set, i.e. user preferences are also concepts: $P \subseteq C$.
If the type of user action included in the user's usage history demands it (like a searchaction), the set of images presented to the user prior or after to that action is also preserved. These constitute the history images associated to the specific user profile and user preferences are derived directly from them. Each history image is represented as a cardinal set on the set of concepts that are related to it and preferences are mined by applying clustering algorithms on it. Most clustering methods belong to either partitioning or hierarchical, however the former require the number of clusters as input and thus are inapplicable. The proposed approach may be decomposed into the following steps:
Perform a sample clustering of concepts in order to determine the count of distinct preferences that a history image is related to, according to the following steps:

1. Turn each available concept into a singleton, i.e. into a cluster k of its own.
2. For each pair of clusters $k1$, $k2$ calculate their distance $d(k1,k2)$.
3. Merge the pair of clusters that have the smallest distance $d(k1,k2)$.
4. Continue at step 2, unless termination criteria are met; termination criterion most commonly used is a threshold for the value of $d\ (k1, k2)$.

Find the user preferences that are related to each cluster.
Aggregate the findings for each cluster to acquire an overall result for each d.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Still, this clustering method creates only crisp clusters and does not allow for overlapping among the detected clusters. The set of preferences that correspond to a history image is the set of preferences that belong to any of the detected odd clusters of concepts.

C. SEARCH RESULT CLUSTERING ALGORITHM

Presentation of search results is perhaps one of the most important factors in the acceptance and popularity of an image retrieval system. We characterize common visualization schemes for image search as follows. Note that the retrieval results are evaluated on a certain number of top-ranked images and preferences and not on the entire ranked list. Algorithm and its specific steps are explained below.

Relevance-Ordered

The most popular way to present search results is relevance ordered, as adopted by Google and Yahoo! for their image search engines. Results are ordered by some numeric measure of relevance to the query.

Image retrieval from Search engine

for image n ∈C (C is the image collection)
for image m ∈C except n
calculate distance d between n and m
ifd <T (T is pre-specified threshold)
putm in the cluster c_n
if the size of cluster $|c_n|$ >S (S is the pre-specified threshold)
break and continue from step 1 with next n
otherwise, continue from step 2 with next m
end

Clustered

Clustering of images by their metadata or visual content has been an active research topic for several years. Clustering of search results, besides being an intuitive and desirable form of presentation, has also been used to improve retrieval performance. The proposed SRC-based RF mechanism employs an effective search result clustering (SRC) algorithm to obtain salient phrases, based on which we could construct an accurate and low-dimensional textual space for the resulting Web images. Given the textual space, we could integrate RF into Web image retrieval in a practical way.

Search result based on user preferences

for image index l ∈L (L is the initial ranked list based on user preferences)
findn=r
for index c ∈c_n
placec just after l
if|ˆL| = E (E is pre-specified length of list to be evaluated)
break and continue from step 1
otherwise, continue from step 3
end

The proposed mechanism shows advantage over traditional relevance feedback methods in the following two aspects. On the one hand, our relevance feedback scheme could catch and reflect user's search intension precisely, for the noisy terms would be exempted from the term list with the aid of clustering, thus, the usability of RF in textual space for Web image retrieval is guaranteed. On the other hand, with the exemption of noisy term, the computation with regards to the low-dimensioned textual space is feasible; therefore, the issues of scalability and efficiency for Web image retrieval are addressed. Experimental results on a database consisting of web images show that the proposed mechanism is wieldy, scalable and effective.

VI. CONCLUSION

The use of visual information in web-based image retrieval is challenging. We developed a method that uses nearly identical visual knowledge obtained from data mining pre-processing in the re-ranking of retrieval results. Refinement of our approach may be possible in the following directions: the use of more sophisticated visual features,

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

the use of collection dependent metrics for comparing images, developing more advanced clustering techniques, and making the threshold values in the data mining process adaptive.

In summary, the main characteristic of SRC is to well-organize and optimize the retrieval quality of interactive CBIR. On one hand, the navigation patterns derived from the users' long term browsing behaviors are used as a good support for minimizing the number of user feedbacks. On the other hand, the projected algorithm SRC Search performs the navigation- pattern-based search to match the user's intention by merging the preferences. As a result, traditional problems such as visual diversity and exploration convergence are solved.

The system has proposed an automatic algorithm for harvesting the web and gathering hundreds of images of a given query class. Through quantitative evaluation has shown that the proposed algorithm performing Google image search and recent techniques that rely on manual intervention. Recent work addresses the only working with few images that are down-loaded and shown the result and we are shown in real time environment and the work would be interesting. Meanwhile, we do hope that the quest for a robust and reliable image understanding technology will continue. Together these will constitute the agenda for future research.

REFERENCES

[1] A.Farhadi, I.Endres,andD.Hoiem, Attribute-centric recognition cross category generalization, in Proc. CVPR, pp.110-112,2010.
[2] Y. Kuo, W. Cheng, H. Lin, and W. Hsu, Unsupervised semantic feature discovery for image object retrieval and tag refinement, IEEE Trans on Multimedia, vol.14, pp.10791090, 2012.
[3] Cutting D. R., Karger D. R., and Pederson J. O. Constant Interaction-Time Scatter/Gather Browsing of Very Large Document Collections. In Proceedings of the 16th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval (SIGIR'93), pages 125-135, Pittsburgh, PA, 1993.
[4] J. Cui, F. Wen, and X. Tang. Intentsearch: Interactive on-line image search re-ranking. In Proc. ACM Multimedia. ACM, 2008.
[5] J. Cui, F. Wen, and X. Tang. Real time google and live image search re-ranking. In Proc. ACM Multimedia, 2008.
[6] Hearst M. A., Pedersen J. O. Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results. In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'96), Zurich, June 1996.
[7] J. Cui, F. Wen, and X. Tang, Real time Google and live image search re-ranking, in Proc. ACM Multimedia, pp.101-104, 2008.
[8] Z. Chen, W.Y. Liu et al. "iFind: a Web Image Search Engine", *SIGIR*. pp.450, 2001.
[9] J. Jeon, R. Manmatha. "Automatic Image Annotation of News Images with Large Vocabularies and Low Quality Training Data", ACM Multimedia, 2004
[10] Leouski A. V. and Croft W. B. An Evaluation of Techniques for Clustering Search Results. Technical Report IR-76, Department of Computer Science, University of Massachusetts, Amherst, 1996.
[11] Leuski A. and Allan J. Improving Interactive Retrieval by Combining Ranked List and Clustering. Proceedings of RIAO, College de France, pp. 665-681, 2000.
[12]Y. J. Du, H. M. Li, "Strategy for mining association rules for web pages based on formal concept analysis", Applied Soft Computing, Vol. 10, No. 3, 2010, pp. 772-783.
[13] V. Nebot, R. Berlanga, "Mining association rules from semantic web data", Proceedings of Trends in Applied Intelligent Systems, SPRINGER-VERLAG BERLIN, June 01-04, 2010, pp. 504-513.
[14] N. Rasiwasia, P. J. Moreno, and N. Vasconcelos, Bridging the gap: Query by semantic example, IEEE Trans. on Multimedia, 2007.
[15] N. Kumar, P. Belhumeur, and S. Nayar. Facetracer: A search engine for large collections of images with faces. ECCV,2008.
[16]Zamir O., Etzioni O. Grouper: A Dynamic Clustering Interface to Web Search Results. In Proceedings of the Eighth International World Wide Web Conference (WWW8), Toronto, Canada, May 1999.
[17] Zamir O., Etzioni O. Web Document Clustering: A Feasibility Demonstration, Proceedings of the 19th International ACM SIGIR Conference on Research and Development of Information Retrieval (SIGIR'98), 46-54, 1998.
[18] G. Carneiro, and N. Vasconcelos, "A Database Centric View of Semantic Image Annotation and Retrieval", *SIGIR*, pp.559-566, 2005.
[19] E. Chang, G. Kingshy et al. "CBSA: Content-based Soft Annotation for Multimodal Image Retrieval Using Bayes Point Machines". IEEE Trans. on CSVT, pp.13(1):26-38, 2003.
[20] B.T. Li, K. Goh, E. Chang. "Confidence-based Dynamic Ensemble for Image Annotation and Semantics Discovery", ACM Multimedia, pp.195-206, 2003.
[21] J. Jeon, V. Lavrenko et al. "Automatic Image Annotation and Retrieval Using Cross-Media Relevance Models". SIGIR, pp.119- 126, 2003.
[22] J. Jeon, R. Manmatha. "Automatic Image Annotation of News Images with Large Vocabularies and Low Quality Training Data", ACM Multimedia, 2004
[23] I.T. Joliffe. "Principal Component Analysis", New York: Springer-Verlag, 1986.
[24] V. Lavrenko, R. Manmatha et al, "A Model for Learning the Semantics of Pictures", NIPS, pp.553-560, 2003.
[25] J. Li, G. Wang et al. "OPTIMOL: automatic Object Picture collection via Incremental MOdel Learning". CVPR,2007.
[26]M.D. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Steele, and P. Yanker, "Query by Image and Video Content: The QBI System," Calculater, vol. 28, no. 9, pp. 23-32, Sept. 1995.