Chapter 9

Building Multi-Modal Relational Graphs for Multimedia Retrieval

Jyh-Ren Shieh
National Taiwan University, Taiwan

Ching-Yung Lin
IBM T. J. Watson Research Center, USA

Shun-Xuan Wang
National Taiwan University, Taiwan

Ja-Ling Wu
National Taiwan University, Taiwan

ABSTRACT

The abundance of Web 2.0 social media in various media formats calls for integration that takes into account tags associated with these resources. The authors present a new approach to multi-modal media search, based on novel related-tag graphs, in which a query is a resource in one modality, such as an image, and the results are semantically similar resources in various modalities, for instance text and video. Thus the use of resource tagging enables the use of multi-modal results and multi-modal queries, a marked departure from the traditional text-based search paradigm. Tag relation graphs are built based on multi-partite networks of existing Web 2.0 social media such as Flickr and Wikipedia. These multi-partite linkage networks (contributor-tag, tag-category, and tag-tag) are extracted from Wikipedia to construct relational tag graphs. In fusing these networks, the authors propose incorporating contributor-category networks to model contributor’s specialization; it is shown that this step significantly enhances the accuracy of the inferred relatedness of the term-semantic graphs. Experiments based on 200 TREC-5 ad-hoc topics show that the algorithms outperform existing approaches. In addition, user studies demonstrate the superiority of this visualization system and its usefulness in the real world.

INTRODUCTION

The last few years we have witnessed the phenomenal success of Web 2.0, which has enabled users to create and exchange self-organized resources on the web, resulting in a huge amount of resources in “folksonomy” systems such as Flickr, YouTube, and De.li.cious. As of October 2009, Flickr for example hosted more than four billion images with manual, user-annotated tags. Tagging functions are widely available in Web 2.0 applications. A tag is a non-hierarchical keyword or term assigned to

DOI: 10.4018/978-1-4666-2940-0.ch009
a piece of information such as an Internet bookmark, an image, or a video. This kind of metadata helps describing an item and allows it to be found again by browsing or searching. Tags are typically created by the media creator or by the viewers in a more or less discretionary manner, but are assumed to be semantically related to their target resources. Image tags in Flickr, for example, help users to understand the background, location, and people in each specific image. Intuitively, if two multimedia resources are tagged with similar tags, a semantic relatedness between them can be established with higher confidence, regardless of their media types. As such, an accurate measure of semantic relatedness among tags will go a long way to realizing the holy grail of semantically relevant multimedia search.

Content-based image retrieval (CBIR) techniques involve understanding the image content directly and automatically annotating these images with keyword terms. However, due to the semantic gap between low-level features and high-level concepts, visually similar images are likely to be judged as semantically unrelated, further associating unrelated concepts with the images. Moreover, errors introduced in annotation mapping or in central topic-cohesive tag detection may result in false concepts being assigned to the resources, which leads to mistaken query results.

In addition to finding an effective way to retrieve semantic related multimedia, we also discovered that the query itself need not be limited to a single modality. Tagging is widely used in Web 2.0 systems, and where a resource is represented in the uniform resource model by its semantically-rich tags, the modality it belongs to is ignored. Therefore, the resources relevant to specific query terms can be determined using tags. Moreover, users can use a single keyword to search for resources from different modalities. Significantly, the queries themselves may also be in various forms, such as text, images, and video. Suppose, for instance, that a user comes across an image on Flickr and is eager to find more information on the subject, such as relevant documents on De.li.ci.ous or videos on Youtube. Unfortunately, such a search is not trivial, as most current systems do not support cross references to related materials and media. In pursuit of this goal, our approach is to create ways to support the discovery of relevant connections across various media as shown in Figure 1. However, establishing connections between related media using naive tag similarity matching can result in either inappropriate relationships or poor search performance. These problems have made finding semantically related tags one of the most challenging issues in Web 2.0 search.

The conventional approach for finding related tags involves the extraction of co-occurring key terms from retrieved documents that are highly ranked. Such approaches are referred to as document-based approaches. To extract high-ranking terms, i.e., tags for media, these approaches must ensure that the extracted terms are representative and that word boundaries are correctly identified (Vectomova et al., 2006). Another problem is that the resultant documents might not all be relevant to the queries. In addition, document-based approaches cannot identify key terms that are highly semantically related, but do not frequently co-occur in documents (Xu, 1996). A variation of the document-based approach is to consider the hyperlink graphs of terms (Gracia et al., 2008).

The second approach is log-based term suggestion, where stored logs of previous users’ queries are investigated. Beeferman and Berger (2000) proposed a query clustering method based on “click-through” data, each record of which consists of a user’s query to the search engine and the URLs within the list provided by the search engine that the user actually visited. One can cluster queries with similar clicked URLs by treating click-through data set as a bipartite graph and identifying the mapping between queries and clicked URLs (Ma et al., 2008). The main problem
Related Content

IT Management in Small and Medium-Sized Enterprises
www.irma-international.org/chapter/management-small-medium-sized-enterprises/17484/

Efficient Imbalanced Multimedia Concept Retrieval by Deep Learning on Spark Clusters
Yilin Yan, Min Chen, Saad Sadiq and Mei-Ling Shyu (2017). International Journal of Multimedia Data Engineering and Management (pp. 1-20).
www.irma-international.org/article/efficient-imbalanced-multimedia-concept-retrieval-by-deep-learning-on-spark-clusters/176638/

Multimodal Information Integration and Fusion for Histology Image Classification
Tao Meng, Mei-Ling Shyu and Lin Lin (2011). International Journal of Multimedia Data Engineering and Management (pp. 54-70).
www.irma-international.org/article/multimodal-information-integration-fusion-histology/54462/

QoS Routing for Multimedia Communication over Wireless Mobile Ad Hoc Networks: A Survey
www.irma-international.org/article/qos-routing-for-multimedia-communication-over-wireless-mobile-ad-hoc-networks/176640/

Comparison of Image Decompositions Through Inverse Difference and Laplacian Pyramids
www.irma-international.org/article/comparison-of-image-decompositions-through-inverse-difference-and-laplacian-pyramids/124243/