

## Multimedia Systems

### Editorial - Social media mining and knowledge discovery

Jinhui Tang<sup>(1)</sup>, Dacheng Tao<sup>(2)</sup>, Guo-Jun Qi<sup>(3)</sup>, Benoit Huet<sup>(4)</sup>

(1) Nanjing University of Science and Technology, Nanjing, China

(2) University of Technology, Sydney, Australia

(3) University of Central Florida, Orlando, FL, USA

(4) EURECOM, Sophia-Antipolis, France

With the rapid advances of Internet and Web 2.0, social networking and social media become more and more popular in humans' daily lives. The ubiquitous nature of web-enabled devices, including desktops, laptops, tablets, and mobile phones, enables users to participate and interact with each other in various web communities, including photo and video sharing platforms, forums, newsgroups, blogs, micro-blogs, bookmarking services, and location-based services. The rapidly evolving social networks provide a platform for communication, information sharing, and collaboration among friends, colleagues, alumnus, business partners, and many other social relations. To be accompanied by, increasingly rich and massive heterogeneous media data have been generated by the users, such as images, videos, audios, tweets, tags, categories, titles, geo-locations, comments, and viewer ratings, which offer an unprecedented opportunity for studying novel theories and technologies for social media analysis and mining. While researchers from multidisciplinary areas have proposed intelligent methods for processing social media data and employing such rich multi-modality data for various applications, it is of high interest to discover potentially important knowledge by social media mining in this nascent field.

Recently, more and more research efforts have been dedicated to the aforementioned challenges and opportunities. This special issue includes five papers focusing on different aspects of social media mining and knowledge discovery.

With the popularity of social media applications, large amounts of social images associated with user tagging information are available, which can be leveraged to boost image retrieval performance. In "Sparse Semantic Metric Learning for Image Retrieval", Liu et al. propose a sparse semantic metric learning method by discovering knowledge from these social media resources, and apply the learned metric to search relevant images for users. Different from traditional metric learning approaches that use similar or dissimilar constraints over a homogeneous visual space, the proposed method exploits heterogeneous information from the visual features and the tagging information of images, and formulates the learning problem as a sparse constrained one. Extensive experiments were conducted on a real-world dataset to validate the effectiveness of the proposed approach.

In most cases, visual information can be regarded as an enhanced content of the textual document. In "Relative Image Similarity Learning with Contextual Information for Internet Crossmedia Retrieval", to make image-to-image similarity being more consistent with document-to-document similarity, Jiang et al. propose a method to learn image similarities according to the relations of the accompanied textual documents. More specifically, instead of using the static quantitative relations, rank-based learning procedure by employing structural SVM is adopted, and the ranking structure is established by comparing the relative relations of textual information. The proposed method can be used not only for the image-to-image retrieval, but also for cross-modality multimedia retrieval.

Integrating multimodal features for visual object recognition is a challenging problem. In paper "Grassmann Multimodal Implicit Feature Selection", Song et al. propose a novel multimodal feature selection method, termed Grassmann manifold feature selection. In this approach, by defining a clustering criterion, the multimodal features are transformed into a matrix, and further treated as a point on the Grassmann manifold. To deal with the unavailable features, the authors compute the L2-Hausdorff distance, a metric between different-sized matrices, to obtain the kernel. Based on the kernel, they propose supervised/unsupervised feature selection algorithms to achieve a physically meaningful embedding of the multimodal features.

To understand how the social network users' video sharing behavior, which is largely based on social relationship, impacts users' viewing pattern, Zhang et al. have conducted a long-term measurement with Renren and Youku, the largest online social network and the largest video sharing site in China, respectively. In "Exploring Sharing Patterns for Video Recommendation on YouTube-like Social Media", the authors show that social friends have higher common interest and their sharing behaviors provide guidance to enhance recommended video lists. They first learn the online social networks (OSNs) video sharing patterns. An auto-encoder model is developed to learn the social similarity of different videos in terms of their sharing in OSNs. Then they propose a similarity-based strategy to enhance video recommendation for YouTube-like social media.

Microtexts is a valuable and albeit noisy source to infer collaborative information. As music plays an important role in many human lives, micro-blogs on music-related activities are available in abundance. In "Harvesting Microblogs for Contextual Music Similarity Estimation", Schedl et al. investigate different strategies to estimate music similarity from these data sources. In particular, they first present a framework to extract co-occurrence scores between music artists from micro-blogs and then investigate L2 similarity estimation functions to subsequently derive resemblance scores. They evaluate the approaches on a collection of micro-blogs crawled from Twitter over a period of ten months and compare them to standard tf-idf approaches. They show that collaborative chatter on music can be effectively used to develop music artist similarity measures.