Heterogeneous Data Alignment for Cross-Media Computing

Shikui Wei† ⋆ Yunchao Wei ⋆ Lei Zhang ⋆ Zhenfeng Zhu ⋆ Yao Zhao ⋆
† Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering,
China Three Gorges University, Yichang, 443002, China.
⋆ Institute of information Science, Beijing Jiaotong University, Beijing, 100044, China
{shkwei, 11112065, 10112061, yzhao, zhfzhu}@bjtu.edu.cn

ABSTRACT

Massive data sets are generated everyday, and knowledge spreading trends to involve diverse media types and information sources. The new situation emerges some new challenging research problems: (1) How to bridge the heterogeneity and mine the shared information among cross-view representations? (2) How to build the semantic association among heterogeneous information objects. (3) How to fully explore the complementary information underlying heterogeneous information objects to cooperatively make decision. In fact, the core of these problems is to align heterogeneous data. In this paper, we first give a short introduction about heterogeneous data alignment. And then, two ongoing works in our group about heterogeneous data alignment are discussed, i.e., consistent pattern mining and modality-dependent cross-media retrieval. Finally, we conclude this paper and discuss some potential applications of heterogeneous data alignment.

Categories and Subject Descriptors

H.3.3 [[**Information Storage and Retrieval**]: Information Search and Retrieval

Keywords

Heterogeneity, Alignment, Cross-Media

1. INTRODUCTION

In the past two decades, both the generation and dissemination manners of information have been evolving rapidly. For the information generation, traditional data sets are generally gathered from expensive devices by some specific organizations, such as news agents. But now almost everyone can generate data sets with the increasingly population of cheap and numerous information-sensing mobile devices. On the one hand, the massive explosion of information objects make the network information resources greatly enriched. On the other hand, it is more and more difficult for users to obtain valuable knowledge when facing the massive information resources, resulting in a serious problem of information overload. For the information dissemination, it is becoming cross media properties, which crosses different information sources, diverse modalities, and various characteristics of properties. The information objects from different sources and modalities are mixed together to reflect underlying semantic meanings, we call it cross media. Cross-media computing aims to effectively mine and represent valuable knowledge from huge and heterogeneous data sets. One of key problem for cross-media computing to align heterogeneous data so as to mine semantic association among them. Figure 1 illustrates an example of heterogeneous data alignment. Here, three different media types, e.g., text, voice and image, are employed to describe the same scene. Although these three media types are semantic complementary to describe the scene, we cannot directly connect them due to the heterogeneous characteristics in low-level features. Therefore, it is difficult to fully explore the complementary information to cooperatively aware the world. To address the issue, we have to align the heterogeneous data. In this paper, we will give a short introduction about our ongoing works on heterogeneous data alignment, which include consistent pattern mining and modality-dependent cross-media retrieval.

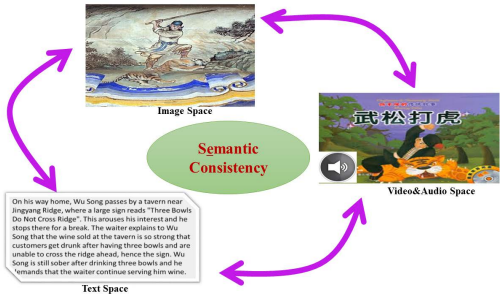

Figure 1: Illustration of heterogeneous data alignment

2. OUR ONGOING WORKS

2.1 Consistent Pattern Mining

The goal of consistent pattern mining is to discover the semantically consistent patterns(SCP) among heterogeneous-view data. In our previous work[2], we propose a three-level framework to align heterogeneous data. Instead of di-

ICIMCS '15, August 19-21, 2015, Zhangjiajie, Hunan, China

DOI: http://dx.doi.org/10.1145/2808492.2808576

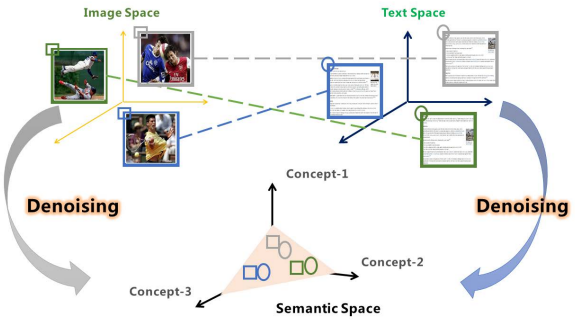

Figure 2: Data denoising for heterogeneous data alignment

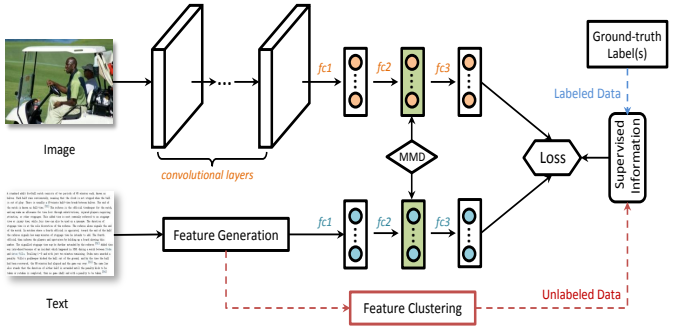

Figure 3: A possible framework for involving unlabeled data for training

rectly mapping multiple views into a shared common space, the proposed framework firstly maps multiple heterogeneous low-level feature spaces into a high-dimensional and redundant feature-isomorphic space. In this way, complementary information underlying different views can be fully captured. Once all the heterogeneous data objects are projected into the feature-isomorphic space, a correlation-based joint feature learning (CJFL) model is proposed to learn a shared semantic subspace. In this step, the redundant and noisy information can be eliminated. However, the existing works, including our work, assume that we have clear training data set. That is, we know the accurate semantic association of the heterogeneous data objects. However, it is less possible for real-world application scenarios to provide clear training data. Therefore, it is necessary to take noise into account when designing consistent pattern mining scheme. In our ongoing work, we will introduce a denoising step during semantic space building process as shown in Figure 2.

2.2 Modality-dependent Cross-media Retrieval

To align heterogeneous data, a commonly used method is to learn a set of project functions for multiple heterogeneous media types. However, in the most cases, all the heterogeneous views are treated equally, which lead to a tradeoff among their respective performances. To address this issue and fully explore complementary information, our previous work [1] proposes a modality-dependent heterogeneous data alignment for cross-media retrieval. Instead of learning a specific set of projecting functions, the proposed approach learns projections according to different cross-media retrieval tasks. That is, the set of projecting functions for Image2Text searching task is different to Text2Image searching task. In this way, the linear regression from a certain modal space and the semantic association among different views can be jointly optimized to achieve the best performance for a specific task. However, both this method and some previous works highly rely on supervised information (i.e., ground-truth label(s) of heterogeneous data) for common representation learning, which will hinder the usage of large-scale unlabeled data. To address this problem, we will design a unified cross-modal framework, in which both labeled data and unlabeled data can be employed for training. The new framework is based on two kinds of deep neural networks and can be trained end-to-end. As shown in Figure 3, we consider two media types, i.e., image and text. We utilize convolutional neural network and fully-connected (fc) neural network to generate feature representations for image and text, respectively. If pairs of image and text are accompanied with ground-truth labels, then the labeled information are employed to optimize the two networks. Otherwise, since the text features are always own more discriminative ability than that of visual features, the clustering results of text features are utilized as the supervised information for network optimization.

3. CONCLUSIONS

In this paper, we first give short discussion about the background and key problems of heterogeneous data alignment. Then two ongoing works, i.e., consistent pattern mining and modality-dependent cross-media retrieval, are described. For the consistent pattern mining, we review our three-level framework for discovering the semantically consistent patterns and point out the direction of future work. Similarly, for the modality-dependent cross-media retrieval, we also review the modality-dependent learning strategy in our previous work and illustrate possible improvement. In addition to cross-media retrieval, the approaches of heterogeneous data alignment can be applied to many other application scenarios, such image annotation, caption generation of image, decision-making, and redundancy removement.

4. ACKNOWLEDGMENTS

This work was supported in part by National Basic Research Program of China (No.2012CB316400), National Natural Science Foundation of China (No.61202241, No.61210006), Program for Changjiang Scholars and Innovative Research Team in University (No.IRT201206), Fundamental Research Funds for the Central Universi-ties (No.2015JBM028), and Hubei Key Laboratory of Intelligent Vision Based Mon-itoring for Hydroelectric Engineering, China Three Gorges University, China (No. 2014KLA03).

5. REFERENCES

[1] Y. Wei, Y. Zhao, Z. Zhu, S. Wei, and F. J. Y. S. Xiao, Yanhui. Modality-dependent cross-media retrieval. *ACM Transactions on Intelligent Systems and Technology*, 2015.

[2] L. Zhang, Y. Zhao, Z. Zhu, S. Wei, and X. Wu. Mining semantically consistent patterns for cross-view data. *IEEE Transactions on Knowledge and Data Engineering*, 26(11):2745–2758, 2014.