

I. Research Overview and Outcome

Introduction

- Documents and records stored by most companies is multi-type
- Multi-type data is a more complicated search task than single-type data
- Searching multi-type data requires the comparison of dissimilar types

Translated Learning

- Translated learning [1] is process of learning in one domain and then using this knowledge in another domain
- The purpose of translated learning is to build a "bridge" to link one feature space (known as the "source space") to another space (known as the "target space") through a translator in order to migrate the knowledge from source to target

- The translated learning model can be represented as follows[1]:

$$C \rightarrow y \rightarrow x$$

Where y represents the features of the data instances x

$$C \rightarrow y_s \rightarrow x_s$$

The training data x_s is represented by features y_s

$$C \rightarrow y_t \rightarrow x_t$$

The test data x_t is represented by features y_t

$$C \rightarrow y_s \rightarrow y_t \rightarrow x_t$$

We create a new chain that connects the features y_s to y_t , and allows use to map features to x_t . The mapping from y_s to y_t is done through the creation of a translator $\phi(y_s, y_t)$

Problem Statement

- Features extracted from dissimilar types can not be compared
- Modeling the relationships between these types is difficult
- Features from dissimilar types must be linked or transformed in order to be compared

Semi-Supervised Learning

- Hidden Markov Random Field (HMRF) [2] is a semi-supervised learning technique that takes known labeled data and pair wise attributes which link objects as either must-link or cannot-link
- Instead of having an underlying Markov chain, hidden Markov random fields have an underlying Markov random field
- The main difference with a hidden Markov model is that neighborhood is not defined in one dimension but within a network
- Objects which have must-links will be heavily weighted to be given the same label; objects which have cannot-links will be heavily weighted to be given different labels

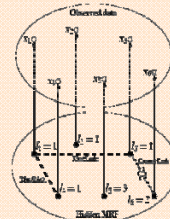


Figure 1. Hidden Markov Random Field [2]

Feature Extraction

- Text, image and video features were extracted. The features between text and multimedia are not comparable.
- Text Features: term frequency-inverse document frequency (tf-idf)
- Image Features: Color histogram, Texture histogram, Blob detection, SIFT points
- Video Features: Extract keyframes and extract features from keyframe images

Future Work

- Modification of HMRF to allow for application of transferred learning to create links
- Develop method to evaluate the effectiveness of multi-type search over single-type search
- Develop a full scale web system to implement the proposed techniques for web based search

Conclusion

- Proposed multi-type search has promise and can be used to create refined search methodologies
- The ability to transfer features between types allows for increased application of traditional search methodologies over non-traditional data sets

References
[1] Wenyuan Dai, Yuqiang Chen, Gui-Rong Xue, Qiang Yang, Yong Yu, "Translated Learning: Transfer Learning across Different Feature Spaces". In Proceedings of NIPS 2008, pp.353-360
[2] S. Basu, M. Bilenko, and R. J. Mooney, "A probabilistic framework for semi-supervised clustering," Seattle, WA, August 2004, pp. 59-68.

II. International Experience



Temple of Heaven - 天坛 - It is amazing to visit a temple that is near 600 years old and is one of the largest in Beijing



Research Team at IBM-CRL



Summer Palace - 颐和园 - Great views, very relaxing, it is easy to see why the emperor would want to spend his summers here

Experience of a lifetime

- See a beautiful foreign country
- Experience new and interesting culture, food and language
- Leave your comfort zone and learn more about your self
- Meet and form collaborations with other researchers from around the world.



Fragrance Hills - 香山 - Great views over western Beijing can be found if you are willing to make the 2 hour hike to the top



Food: The food was delicious, I have never eaten so many different types of food. I tried food that I never would have at home and found it delicious



Tsinghua View: Some days were better than others. The campus was much larger than any college campus I had ever seen

III. Acknowledgement

The material presented in this poster is based upon the work supported by the National Science Foundation under Grant No. OISE-0730065. Additional support was provided by the GAANN Fellowship funded by U.S. Department of Education under Grant No. P200A070543. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the Department of Education.