

Correlation models for Learning from Images and Text

Oksana Yakhnenko

Learning from Multi-Modal data

Challenge

Data comes from multiple sources: images, text, sound, video

Labeled data is expensive, but plenty unlabeled/weakly-labeled data is available (ie. Flickr, Picasa, social network communities)
In image annotation can we relax the need for all regions to be labeled

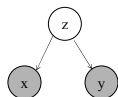
- use only image and caption information



How can we learn and utilize correlations among various data sources?

Two views of modeling correlations

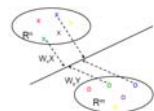
Project multiple views of data onto common low-dimensional latent semantic space



Probabilistic view of correlation

For one view: pLSA, LDA...

For multiple views: MoMLda, CorLDA...



Linear view of correlation

For one view: LSI, PCA

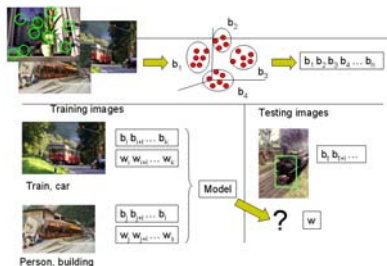
For multiple views: CCA/KCCA

Data

VOC 2007

2500 training images
5000 test images
20 possible words

Images as bags of Visual Words



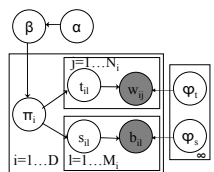
LabelMe

2600 images
50/50 train/test split
~300 possible words

Corel 5K

4500 training images
500 test images
~260 possible words

Multi-Modal Hierarchical Dirichlet Process



Key properties:

Models probability of images and keywords
Assumes infinitely many mixture components
Truncated stick breaking distribution

- Model selection: allows mixture components adapt to the data
- Parameter estimation using variational inference – fast!
- Better performance than one-against-all methods

Kernel Correlation Models

Kernel Multiple Linear Regression

Discriminative version of KCCA

Minimize the reconstruction error directly: $\min_w \left\| \sum_{i=1}^K d_i w_i^T X w_{iy} - Y \right\|^2$

Dual form: allows the use of kernel $\max \rho = \frac{\alpha^T K_x K_y^T \beta}{\sqrt{\alpha^T K_x^2 \alpha} \sqrt{\beta^T K_y \beta}}$

Basis α, β are solutions to generalized eigen-value problem

Kernel Selection for KCCA/KMLR

Images and text can be represented in variety of ways (visual words, graphs, color histograms, segmentation)

Need for efficient kernel selection and kernel combination

Consider convex combination of kernels

$$K = \sum_{i=1}^m \eta_i K_i$$

$$\text{sbj. to } \sum_{i=1}^m \eta_i = 1, \eta_i \geq 0$$

Iterative process for kernel selection:

1. Solve KCCA/KMLR for α, β using fixed η
2. Use gradient ascent to solve for η using α, β

Main Contributions

MoM-HDP

- Circumvent model selection for probabilistic
- Allows model adapt to the data
- Prone to overfitting on the training data as the number of mixture components increases

KMRL

- Allows to incorporate non-linear correlations on the using kernel
- Discriminative version of KCCA – lower caption reconstruction error
- Framework for automatic kernel selection

Work in Progress

- Experimental evaluation on other datasets for image annotation/retrieval
- Experimental evaluation on cross-lingual retrieval
- Experimental evaluation of kernel selection on a variety of image kernels

Selected References

[1] Oksana Yakhnenko and Vasant Honavar. *Multiple label prediction for image annotation with multiple kernel correlation models*. Workshop on Visual Context Learning at CVPR 2009.

[2] Oksana Yakhnenko and Vasant Honavar. *Multi-Modal Hierarchical Dirichlet Process Model for Predicting Image Annotation and Image-Object Label Correspondence*. SIAM International Conference on Data Mining SDM 2009.

Contact Information

Oksana Yakhnenko: oksayakh@cs.iastate.edu

Thesis Advisor: Vasant Honavar