Learning Joint Representation for Community Question Answering with Tri-modal DBM

Baolin Peng[1], Wenge Rong[1,2], Yuanxin Ouyang[1,2], Chao Li[1,2], Zhang Xiong[1,2]
[1]School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[2]Research Institute of Beihang University in Shenzhen, Shenzhen 518057, China
b.peng@cse.buaa.edu.cn, {w.rong, oyyx, licc, xiongz}@buaa.edu.cn

ABSTRACT

One of the main research tasks in Community question answering (CQA) is to find most relevant questions for a given new query, thereby providing useful knowledge for the users. Traditionally used methods such as bag-of-words or latent semantic models consider queries, questions and answers in a same feature space. However, the correlations among queries, questions and answers imply that they lie in different feature spaces. In light of these issues, we proposed a tri-modal deep boltzmann machine (tri-DBM) to extract unified representation for query, question and answer. Experiments on Yahoo! Answers dataset reveal using these unified representation to train a classifier judging semantic matching level between query and question outperforms models using bag-of-words or LSA representation significantly.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information ServicesWeb-based services

General Terms

Algorithms, Experimentation

Keywords

Deep Boltzmann Machine; Query Understanding; Community Question Answering; Semantic Similarity

1. INTRODUCTION

In recent year, community question answering (CQA) sites has become a popular platform for users to share knowledge. One typical CQA based information seeking process is to find similar questions from the past with regard to new queries to obtain useful knowledge for the users. Therefore it is essential to have effective mechanism to measure the similarity between queries and existing questions.

The widely used algorithm is TF-IDF based on bag-of-word representation, which however cannot model semantic similarity when keyword-based matching fails. To tackle this drawback, latent semantic analysis (LSA) is proposed to extract low-dimensional semantic representation by singular value decomposition (SVD) decomposition.

Either bag-of-word or LSA regards queries, questions and answers in a same feature space. However in reality queries tend to be short and normally not a complete sentence, questions are usually longer and typically be one sentence, while answers are the most informative one which genearlly comprise many sentences. As such it is argued that they are in three different modalities.

Inspired by this assumption, a tri-modal deep Boltzmann machine (tri-DBM) is propsed to get a unified representation [2] from three separate feature spaces for queries, questions and answers. Firstly, three separate deep Boltzmann machines (DBM) are employed to model highly non-linear correlations in their own feature space. Afterwards another DBM is used to get a unified representation in a joint space. Experimental study on Yahoo! Answers dataset of "query and question semantic level classification" shows using the unified representation significantly outperform baselines using bag-of-word, LSA or unimodal DBM representation.

2. APPROACH

Restricted Boltzmann machines (RBMs) have been widely used as building block for deep learning algorithms due to its power in modeling distributions over binary-valued data. In this research, we take a variant RBM named Replicated Softmax Model (RSM) for building block since it has been proved to be useful for modeling sparse data [3]. Similar to traditional RBMs, RSM consists of two layer states, i.e., stochastic visible state $\mathbf{v}$ and stochastic hidden state $\mathbf{h}$. Its energy function of state $\mathbf{v}$ and $\mathbf{h}$ is computed as follows:

$$E(\mathbf{v},\mathbf{h};\theta) = -\sum_{i=1}^{m}\sum_{j=1}^{n} W_{ij}v_i h_j - \sum_{i=1}^{m} b_i v_i - M\sum_{j=1}^{n} c_j h_j \quad (1)$$

where $\theta = (W, b, c)$ are model parameters to be learned and M is total number of words occurred in a query or question.

Figure 1 depicts the architecture of triDBM. Firstly, three separate DBMs are used to model highly abstracted representation for queries, questions and answers respectively. Secondly, the three DBMs are combined by adding an additional layer of another DBM to model non-linear correlation among different input modals. The model is firstly trained with 1-step Contrastive Divergence [1]. Afterwards the upper layer will become the joint representation which can be used for classification or retrieval tasks.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577341.

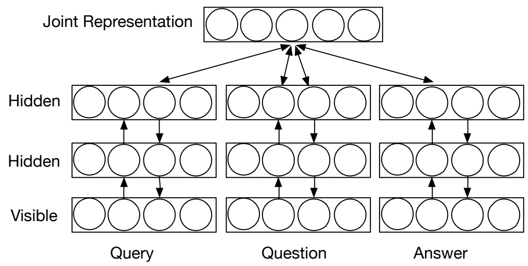

Figure 1: Architecture of Tri-DBM

3. EXPERIMENTAL SETUP

Dataset: The dataset used in this experimental study is from a public dataset, Yahoo! Answers query to questions dataset, which has 12850 $< query, question, anserw >$ triples. Each item is labelled as 1, 2 or 3 to indicate the semantic relevance level between query and question.

Metrics: *accuracy* is used and can be defined as follows:

$$accuracy = \frac{\sum_i^n 1\{p_i == g_i\}}{\#test} \tag{2}$$

Procedure: To create tri-modal DBM, we utilize a #dict-500-128 architecture for hidden and visible layer. #dict means dictionary size which in query, question, and answer modal are 1320, 3792 and 6210 respectively. RBM used in joint representation layer has 384-128 architecture. Each layer is greedily pretrained for 300 passes. After getting joint representation, it will be fed into muti-class logistic classifier to predict semantic similarity level of query and question. The dataset is randomly split into training set (70%) and testset(30%). Each method are conducted ten times and the average accuracy is considered as final result.

4. RESULTS AND DISCUSSION

LDA and pLSA models do not generally outperforms LSA in preliminary experiments. Therefore, we utilize LSA as baseline approach. From figure 2, we found that DBM can effectively capture semantic meaning. Questions or answers with same meaning tend to be close to each other while LSA approach map them almost in a same place.

Table 1 summarized the experimental results of proposed method against some baselines. Row 1 is the method with randomly assigned similarity values. Row 2 is the term vectors based method. Rows 3 to 5 present results of LSA representation. uniLSA means mapping Row 2 representation into a lower dimension and then use it for classification. Row 4 stands for mapping query, question and answer into three separate lower dimension and concatenate them for classification. Row 5 expresses further mapping above concatenated representation into a lower dimension to model correlation between input modals using. Row 6 to 8 show results with DBMs approach. One reason uniDBM or triDBM without joint representation layer (WJRL) lag behind LSAs may be that LSAs can maintain original information to some degree since it is a shallow model and can only capture pairwise semantic similarity while DBM can model highly non-linear correlation. On the contrary, just concatenating their lower dimension representation could hurt performance. However, as row 8 shows, when using another DBM to model these separate representations, triDBM can significantly boost the overall performance.

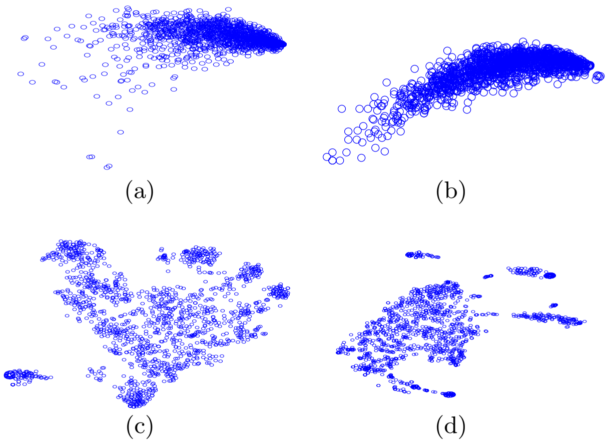

Figure 2: 2(a), 2(b) illustrate question and answer 2-dimensional representation produced by LSA; 2(c),2(d) depict question and answer 2-dimensional representation introduced by tri-DBM without joint representation layer.

Table 1: Accuracy Summary over Different Methods

Method	Accuracy	Method	Accuracy
Random	33.33	triLSA	43.90
Bag-of-Words	51.22	uniDBM	45.53
uniLSA	50.44	triDBM WJRL	44.72
triLSA WJRL	52.85	triDBM	**64.23**

5. CONCLUSION

In this work, we proposed a tri-modal DBM for predicting query and question similarity by extracting a unified representations among query, question and answer into a joint feature space. We further compare these unified representations against baselines which utilize bag-of-word or LSA representation as features in a discriminative task using data from Yahoo! Answers. The results revealed that tri-modal DBM can captures semantic relationship among query, question and answers. Taking these unified representation as feature can significantly improve the overall performance.

6. ACKNOWLEDGEMENT

This work was partially supported by the National High Technology Research and Development Program of China (No. 2011AA010502) and the National Natural Science Foundation of China (No. 61103095).

7. REFERENCES

[1] G. E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, 2002.

[2] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng. Multimodal deep learning. In *ICML*, pages 689–696, 2011.

[3] R. Salakhutdinov and G. E. Hinton. Replicated softmax: an undirected topic model. In *NIPS*, pages 1607–1614, 2009.