International Journal of u-and e-Service, Science and Technology
Vol.7, No.4 (2014), pp.11-18
http://dx.doi.org/10.14257/ijunnesst.2014.7.4.02

In Search of Synergies between Policy-Based Systems Management and Deep Boltzmann Machines Models for E-commerce

Li Min[1], Liu Wei[1] and Xichun Guo[2]

[1]*Department of Comupter and Information Enginer, Harbin University of commerce*
[2]*Department of logistics management, Harbin Railway Technical College*
E-mail: lm81612@163.com

Abstract

This paper discusses the motivations and principles of Deep Boltzmann Machines regarding learning algorithms for deep architectures. Policy-based systems management (PBM) and Deep Boltzmann Machines (DBM) are two of the many techniques available for artificial intelligence (AI), each having specific benefits and limitations, and thus different applicability; choosing the most appropriate technique is the first of many challenges faced by the developer. The discussion forms a backdrop for a detailed evaluation of the two techniques, in which the concepts underpinning each of PBM and EBM are reviewed and placed into context with each other as well as with the other popular techniques for AI. After considering the operation and suitability of the techniques in isolation, the focus shifts to look at how PBM and DBM could be combined in complementary ways to achieve more sophisticated and versatile AI systems for E-commerce..

Keywords: *Deep learning; Deep Boltzmann Machines; Policy-based systems management; E-commerce*

1. Introduction

Cognitive radio is presented base on software radio for scarcity and underutilization of spectrum resource.

Theoretical results suggest that in order to learn the kind of complicated functions which can represent high-level abstractions (*e.g.*, invision, language, and other AI-level tasks), people may need deep architectures. Deep architectures are composed of multiple levels of non-linear operations, such as in neural nets with many hidden layers or in complicated propositional formulae re-using many sub-formulae. Searching the parameter space of deep architectures is a difficult task, but recently learning algorithms such as RBM and DBN have been proposed to tackle this problem with notable success, beating the state-of-the-art in certain areas.

Policy-based systems management (PBM)

To differentiate between levels of importance we use the labels “High Importance”, “Normal Importance”, and “Best Effort” along with an importance multiplier that indicates how much more important one class is than another. A class that is “High Importance” should have priority over classes that are “Normal Importance” or “Best Effort”. In a consolidated enterprise system, this “High Importance” label could, for example, be applied to a class of work representing an OLTP-like order entry department as this work is directly revenue generating, or to a class of work corresponding to queries entered by the company CEO and thus considered most important. The “Normal Importance” label would likely apply to all

ISSN: 2005-4246 IJUNNEST

other business related workloads such as inventory reports for a procurement department that are needed during business hours, but have a lower priority than business units directly affecting revenue. Finally, the "Best Effort" label would be applied to classes of transactions that do not have any strict deadline, such as background DBMS maintenance work, or after hours reporting queries.

We incorporate the concept of an importance multiplier into the economic model to differentiate among the levels of importance. The multipliers directly affect the amount of wealth a class has to acquire resources. For example, an agent with a "High" degree of importance may have a multiplier of 3.0, while an agent with a "Best Effort" degree of importance may have a multiplier of only 1.0. By adjusting these multipliers, we affect the amount that one class is more important than another by altering their ability to outbid other classes for resources. In the following experiments we look at the impact of a range of values for these multipliers.

The definition of importance describes the differences in entitlements and abilities between a high-priority class and a low-priority class. We base our definition on the entitlement one class has to resources as compared to other classes. Three definitions of importance are defined by Boughton [1]:

Non-Preemptive: a minimum amount of resources are guaranteed to all workloads.

Preemptive: one workload may appropriate all resources, preempting the execution of others.

Preemptive Except High Importance: resources are allocated for the high priority workloads and a preemptive model is used for the remaining resources.

A Deep Boltzmann Machine is described for learning a generative model of data that consists of multiple and diverse input modalities. It can be used to extract a unified representation that fuses modalities together. We find that this representation is useful for classification and information retrieval tasks. The model works by learning a probability density over the space of multimodal inputs. It uses states of latent variables as representations of the input. The model can extract this representation even some modalities are absent by sampling from the conditional distribution over them and filling them in. Our experimental results on bi-modal data consisting of images and texts show that the Multimodal DBM can learn a good generative model of the joint space of image and text inputs that are useful for information retrieval from both unimodal and multimodal queries. We further demonstrate that this model significantly outperforms SVMs (Support Vector Machines) and LDA (Linear Discriminant Analysis) on discriminative tasks.

2. BackgrounD: RBMS and their Generalizations

Restricted Boltzmann Machines (RBMs) have been used effectively in modeling distributions over binary-valued data. Recent work on Boltzmann machine models and their generalizations to exponential family distributions have allowed these models to be successfully used in many application domains.

We propose multimodal Deep Boltzmann Machine (DBM) model that satisfies the above desiderata. DBMs are undirected graphical models with bipartite connections between adjacent layers of hidden units [2]. The key idea is to learn a joint density model over the space of multimodal inputs. Then Missing modalities can be filled in sampling from the conditional distributions over them given the observed ones. For example, we use a large collection of user-tagged images to learn a joint distribution over images and texts P(Vm, Vt

|ϕ) . By drawing samples from P(Vt | Vm,ϕ) and from P(Vm | Vt,ϕ) we can fill-in missing data, thereby doing image annotation and image retrieval respectively.

There have been several approaches to learning from multimodal data. In particular, Huiskes et al. [3] showed that using captions, or tags, in addition to standard low-level image features significantly improves classification accuracy of Support Vector Machines (SVM) and Linear Discriminant Analysis (LDA) models. A similar approach of Guillaumin et al. [4], based on multiple kernel learning framework, further demonstrated that an additional text modality can improve the accuracy of SVMs on various object recognition tasks. However, all of these approaches are discriminative by nature and cannot make use of large amounts of unlabeled data or deal easily with missing input modalities.

On the generative side, Xing et al. [5] used dual-wing harmoniums to build a joint model of images and texts, which can be viewed as a linear RBM model with Gaussian hidden units together with Gaussian and Poisson visible units. However, various data modalities will typically have different statistical properties which make it difficult to model them using shallow models. Most similar to our work is the recent approach of Ngiam et al. [6] that used a deep autoencoder for speech and vision fusion. However, there are several crucial differences. First, in this work we focus on integrating different data modalities together: sparse word count vectors and real-valued dense image features. Second, we develop a Deep Boltzmann Machine for autoencoder generative model as opposed to unrolling the network and fine-tuning. While both approaches have lead to interesting results in several domains. Using a generative model is important for applications we consider in this paper, because it is a method that allows our model to naturally handle missing data.

3. Multimodal Deep Boltzmann Machine

3.1. Salient Features

A Multimodal DBM can be viewed as a composition of unimodal undirected pathways. Each pathway can be pretrained separately in a completely unsupervised fashion, which allows us to leverage a large supply of unlabeled data. Any number of pathways each with any number of layers could potentially be used. The type of the lower-level RBMs in each pathway could be different, accounting for different input distributions, as long as the final hidden representations at the end of each pathway are of the same type.

The intuition behind our model is as follows. Each data modality has different statistical properties which make it difficult for a single hidden layer model (such as Figure 1a) directly to find correlations across modalities. In our model, this difference is bridged by putting layers of hidden units between the modalities. The idea is illustrated in Figure 1c. Compared to the simple RBM (Fig. 1a), where the hidden layer h directly models the distribution over Vt and Vm, the first layer of hidden units hm(1) in a DBM has an easier task to perform - that of modeling the distribution over Vm and hm(2). Each layer of hidden units in the DBM contributes a small part to the overall task of modeling the distribution over Vm and Vt. In the process, each layer learns successively higher-level representations and removes modality-specific correlations. Therefore, the middle layer in the network can be seen as a (relatively) "modality-free" representation of the input opposed to the input layers which were "modality-full".

Another way of using a deep model to combine multimodal inputs is to use a Multimodal Deep Belief Network (DBN) (Fig. 1b) which consists of an RBM followed by directed belief networks leading out to each modality. We emphasize that there is an important distinction between this model and the DBM model of Figure 1c. In a DBN model the responsibility of

the multimodal modeling falls entirely on the joint layer. On the other hand, in the DBM, this responsibility is spread out over the entire network. The modality fusion process is distributed across all hidden units in all layers. From the generative perspective, states of low-level hidden units in one pathway can influence the states of hidden units in other pathways through the higher-level layers, which is not the case for DBNs.

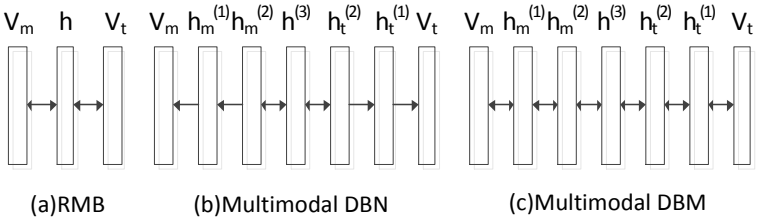

Figure 1. Different Ways of Combining Multimodal Inputs

3.2. Modeling Tasks

Generating Missing Modalities: As argued in the introduction, many real-world applications will often have one or more modalities missing. The Multimodal DBM can be used to generate such missing data modalities by clamping the observed modalities at the inputs and sampling the hidden modalities from the conditional distribution by running the standard alternating Gibbs sampler [2]. For example, considering generating text conditioned on a given image Vm, the observed modality Vm is clamped at the inputs and all hidden units are initialized randomly. P(Vt | Vm) is a multinomial distribution over the vocabulary. Alternating Gibbs sampling can be used to sample words from P(Vt | Vm).

Inferring Joint Representations: The model can also be used to generate a fused representation that multiple data modalities. This fused representation is inferred by clamping the observed modalities and doing alternating Gibbs sampling to sample from P(h(3) | Vm, Vt) (if both modalities are present) or from P(h(3) | Vm,) (if text is missing). A faster alternative, which we adopt in our experimental results, is using variational inference to approximate posterior Q(h(3) | Vm, Vt) or Q(h(3) | Vm,). The activation probabilities of hidden units h(3) constitute the joint representation of the inputs.

Then this representation can be used to do information retrieval for multimodal or unimodal queries. Each data point at the database (whether missing some modalities or not) can be mapped to this latent space. Queries can also be mapped to this space and an appropriate distance metric can be used to retrieve results that are close to the query.

Discriminative Tasks: Classifiers such as SVMs can be trained with these fused representations as inputs. Alternatively, the model can be used to initialize a feed forward network which then can be finetuned [2]. In our experiments, logistic regression was used to classify the fused representations.

Unlike finetuning, this ensures that all learned representations that we compare (DBNs, DBMs and Deep Autoencoders) use the same discriminative model.

4. In Search of Synergies for E-Commerce

4.1. Dataset and Feature Extraction

The MIR Flickr Data set [7] was used in our experiments. The data set consists of 1 million images retrieved from the social photography website Flickr along with their user assigned

tags. Among the 1 million images, 25,000 have been annotated for 24 topics including object categories such as, bird, tree, people and scene categories, such as indoor, sky and night. A stricter labeling was done for 14 of these classes where an image was annotated with a category only if that category was salient. This leads to a total of 38 classes where each image may belong to several classes. The unlabeled 975,000 images were used only for pretraining. We use 15,000 images for training and 10,000 for testing, following Huiskes et al. [3]. Mean Average Precision (MAP) is used as the performance metric. Results are averaged over 5 random splits of training and testing sets. Each text input was represented using a vocabulary of the 2000 most frequent tags. The average number of tags associated with an image is 5.15 with a standard deviation of 5.13. There are 128,501 images which do not have any tags, out of which 4,551 are in the labeled set. Hence about 18% of the labeled data has images but is missing text. Images were represented by 3857-dimensional features, that were extracted by concatenating Pyramid Histogram of Words (PHOW) features [8], Gist [9] and MPEG-7 descriptors [10] (EHD, HTD, CSD, CLD, SCD). Each dimension was mean-centered and normalized to unit variance. PHOW features are bags of image words obtained by extracting dense SIFT features over multiple scales and clustering them. We used publicly available code ([11, 12]) for extracting these features.

4.2. Model

The image pathway consists of a Gaussian RBM with 3857 visible units followed by 2 layers of 1024 hidden units. The text pathway consists of a Replicated Softmax Model with 2000 visible units followed by 2 layers of 1024 hidden units. The joint layer contains 2048 hidden units. Each layer of weights was pretrained using performed using Contrastive Divergence (PCD) for initializing the DBM model. When learning the DBM model, all word count vectors were scaled so that they sum to 5. This avoids running separate Markov chains for each word count to get the model distribution's sufficient statistics.

Each pathway was pretrained using a stack of modified RBMs. Each Gaussian unit has unit variance that was kept fixed. For discriminative tasks, we perform 1-vs-all classification using logistic regression on the joint hidden layer representation. We further split the 15K training set into 10K for training and 5K for validation.

4.3. Classification

Our first set of experiments evaluate the DBM as a discriminative model for multimodal data. For each model that we trained, the fused representation of the data was extracted and feed to a separate logistic regression for each of the 38 topics. The text input layer in the DBM was left unclamped when the text was missing. Tables 1 summarizes the Mean Average Precision (MAP) and precision@50 (precision at top 50 predictions) obtained by different models. Linear Discriminant Analysis (LDA) and Support Vector Machines (SVMs) [2] were trained using the labeled data on concatenated image and text features that did not include SIFT-based features. Hence, to make a fair comparison, our model was first trained using only labeled data with a similar set of features (i.e., excluding our SIFT-based features). We call this model DBM-Lab. Tables. 1 shows that the DBM-Lab model already outperforms its competitor SVM and LDA models. DBM- Lab achieves a MAP of 0.52, compared to 0.47and 0.49, achieved by SVM and LDA models.

Based on PBM in the experiments discussed below， to differentiate between levels of importance we use the labels "High Importance", "Normal Importance", and "Best Effort" along with an importance multiplier that indicates how much more important one

class is than another. In a consolidated the database of images ,this "High Importance" label could, for example, be applied to a class of work representing edge. The "Normal Importance" label would likely apply to Color. The "Best Effort " label would likely apply to all other .

Table 1. Deep Architectures Results

Deep architectures Results		
Model	***MAP***	***Prec@50S***
Random	0.12	0.12
LDA[2]	0.49	0.75
SVM [2]	0.47	0.75
DBM-Lab	0.52	0.79
DBM-Unlab	0.58	0.83
DBN	0.59	0.86
Autoencoder ased on [5]	0.60	0.87
DBM	0.61	0.87

Tables. 2 shows Results Based on Policy-based systems management (PBM) **Tables. 2** shows that the DBM-Lab model already outperforms its competitor **Non-Preemptive** and **Preemptive** models. In **Non-Preemptive** PBM- Lab achieves a Random of 0.13, 0.14 compared to 0.12and 0.12, In **Preemptive** PBM- Lab achieves a Random of 0.22, 0.24 compared to 0.13and 0.14,achieved by DBM and **Non-Preemptive** models.

Table 2. PBM Results

Classification Tasks	*WAY*	*MAP*	*Prec@50S*
Deep architectures	*Random*	*0.12*	*0.12*
	Autoencoder ased on [5]	*0.60*	*0.87*
	DBM	*0.61*	*0.87*
Non-Preemptive	*Random*	*0.13*	*0.14*
	Autoencoder ased on [5]	*0.61*	*0.88*
	DBM	*0.64*	*0.87*
Preemptive	*Random*	*0.22*	*0.24*
	Autoencoder ased on [5]	*0.62*	*0.87*
	DBM	*0.66*	*0.88*

4.4. Quantification

A database of images was created by randomly selecting 5000 image- text pairs from the test set. We randomly selected a disjoint set of 1000 images to be used as queries. Each query contained both image and text modalities. Binary relevance labels were created by assuming that if any of the 38 class labels overlapped between a query and a data point, then that data point is relevant to the query. We base our definition on the entitlement one class has to resources as compared to other classes.Cosine similarity function was used to match queries to data points. The DBM model performs the best among the compared models achieving a MAP of 0.61. The PBM model in Non-Preemptive performs the best among the compared models achieving a MAP of 0.64. The PBM model in Preemptive performs the best among the compared models achieving a MAP of 0.66. Note that even though there is little overlap in terms of text, the model is able to perform well.

5. Conclusion

We proposed a Deep Boltzmann Machine model for learning multimodal data representations. Large amounts of unlabeled data can be effectively utilized by the model. Pathways for each modality can be pretrained independently and "plugged in" together for doing joint training. The model fuses multiple data modalities into a unified representation. This representation captures features that are useful for classification and retrieval. It also works nicely when some of the ways is defective modalities are absent and improves upon models trained on only the observed modalities.

The definition of importance describes the differences in entitlements and abilities between a high-priority class and a low-priority class. We base our definition on the entitlement one class has to resources as compared to other classes. This paper called Policy-based systems management (PBM).

Current training algorithms for deep architectures involves many phases (one per layer, plus a global fine-tuning). This is not very practical in the purely online setting since once we have moved into fine-tuning, we might be trapped in an apparent local minimum. Is PBM possible to come up with a completely online procedure for training deep architectures that preserves an unsupervised component all along? Note that is appealing for this reason.

In Search of Synergies between Policy-Based Systems Management and Deep Boltzmann Machines Models for E-commerce.We can use it for Recommended clothing in E-commerce.

Acknowledgment

This paper is supported by Heilongjiang Society of Vocational Education ducation (GG0667).

References

[1] Boughton, H., Martin, P., Powley, W. and Horman, R. 2006. Workload class importance policy in autonomic database management systems. Proceedings of the 2006 IEEE Workshop on Policies for Distributed Systems and Networks (POLICY 2006), London ON, June 5-7, 2006, 13 – 22.
[2] R. R. Salakhutdinov and G. E. Hinton. Deep Boltzmann machines. In Proceedings of the International Conference on Artificial Intelligence and Statistics, volume 12, 2009.
[3] Mark J. Huiskes, Bart Thomee, and Michael S. Lew. New trends and ideas in visual concept detection: the MIR flickr retrieval evaluation initiative. In Multimedia Information Retrieval, pages 526–536, 2010.
[4] M. Guillaumin, J. Verbeek, and C. Schmid. Multimodal semi-supervised learning for image classification. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 907 –909, june 2010.
[5] Eric P. Xing, Rong Yan, and Alexander G. Hauptmann. Mining associated text and images with dual-wing harmoniums. In UAI, pages 633–641. AUAI Press, 2005.
[6] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y. Ng.Multimodal deep learning. In International Conference on Machine Learning (ICML), Belle-vue, USA, June 2011.
[7] Mark J. Huiskes and Michael S. Lew. The MIR Flickr retrieval evaluation. In MIR '08:Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval, New York, NY, USA, 2008. ACM.
[8] A Bosch, Andrew Zisserman, and X Munoz. Image classification using random forests and ferns. IEEE 11th International Conference on Computer Vision (2007), 23:1–8, 2007.
[9] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision, 42:146–175, 2001.
[10] B.S. Manjunath, J.-R. Ohm, V.V. Vasudevan, and A. Yamada. Color and texture descriptors.Circuits and Systems for Video Technology, IEEE Transactions on, 11(6):703 –717, 2001.
[11] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algo-rithms, 2008.
[12] Muhammet Bastan, Hayati Cam, Ugur Gudukbay, and Ozgur Ulusoy. Bilvideo-7: An mpeg-7-compatible video indexing and retrieval system. IEEE Multimedia, 17:62–73, 2010.

Authors

Li Min, master instructor, hold the post of director of Software Engineering Teaching and Research Section and hold the concurrent post of competition group specialist committee member in Heilongjiang Computer Society. Her main research fields include algorithm design theory, intelligence decision.

Liu Wei, master, Harbin University of Commerce. main research fields include algorithm design theory, intelligence decision.

Xichun Guo, Male, born in Jiangxi, is a lecturer in Harbin Railway Technical College. He received Bachelor of Engineering from East China Jiaotong University in 2007. He is now majoring in Master of Engineering at Harbin Institute of Technology. His research interest is in Image processing program and he has two patents and two published books.