Ontology-Based Deep Restricted Boltzmann Machine

Conference Paper · September 2016

DOI: 10.1007/978-3-319-44403-1_27

Ontology-Based Deep Restricted Boltzmann Machine

Hao Wang(✉), Dejing Dou, and Daniel Lowd

Computer and Information Science, University of Oregon, Eugene, USA
{csehao,dou,lowd}@cs.uoregon.edu

Abstract. Deep neural networks are known for their capabilities for automatic feature learning from data. For this reason, previous research has tended to interpret deep learning techniques as data-driven methods, while few advances have been made from knowledge-driven perspectives. We propose to design a semantic rich deep learning model from a knowledge driven perspective, by introducing formal semantics into deep learning process. We propose ontology-based deep restricted Boltzmann machine (OB-DRBM), in which we use ontology to guide architecture design of deep restricted Boltzmann machines (DRBM), as well as to assist in their training and validation processes. Our model learns a set of related semantic-rich data representations from both formal semantics and data distribution. Representations in this set correspond to concepts at various semantic levels in a domain ontology. We show that our model leads to an improved performance, when compared with conventional deep learning models in classification tasks.

1 Introduction

Deep learning has achieved state of the art performance on many cutting-edge applications, including computer vision [1], speech and phonetic recognition [2], natural language processing [3], multi-task and multi-modal learning [4], and many others. Deep learning is often called *representation learning* [5], which emphasizes its aspect of automatic representation learning from data. Features in learned representations are formulated in a bottom up way, such that higher-level features are defined recursively from lower-level ones. For this reason, previous research tended to interpret deep learning techniques from data-driven perspectives. Few efforts have been made for semantic-rich deep learning methods, especially, for the ones using formal semantics.

In practice, data-driven approaches often carry various limitations. In deep learning, it is often difficult to interpret representations learned from data with accurate high-level semantics [6]. Over-fitting is a prevalent issue in deep neural networks that have a large number of parameters [7]. While we expect a well-trained deep representation to encode a non-local generalization prior over input space, it has often been proved to be sensitive to training data distribution. Poorly distributed data can result in an inferior or even wrong generalization. For similar reasons, deep representations often fail to generalize to examples that

S. Hartmann and H. Ma (Eds.): DEXA 2016, Part I, LNCS 9827, pp. 431–445, 2016.
DOI: 10.1007/978-3-319-44403-1_27

fall outside original training sample domain. For instance, deep neural networks can mis-classify images, when imperceptible perturbation is applied [8]. Or they can interpret images that are completely unrecognizable to humans, with almost full confidence [9].

One prevalent way to solve the afore-mentioned issues in data-driven approaches is to augment machine learning tasks with domain knowledge. Previously, domain knowledge has been applied on a wide range of applications in various forms. However, for those methods with task-dependent domain knowledge, making generalizations to new applications are usually difficult due to their labor-intensive knowledge crafting process. On the other hand, formal semantics, the formal encoding of domain knowledge, has provided a way to systematically encode, share, and reuse knowledge across applications and domains. In practice, formal semantics can support a wide range of key aspects in machine learning, data mining, and artificial intelligence techniques. For instance, formal semantics can help filter out redundant or inconsistent data, and can generate semantic rich results [10]. It can work as a set of prior knowledge or constraints, to help reduce search space and to guide search path [11].

It turns out to be an intriguing question to wonder what roles formal semantics can play in the recent trend of machine learning research, deep learning. Based on previous research, we expect formal semantics to assist in deep learning process from the following perspectives:

- Directing deep learning architecture design, resulting in learning models that better fit with current application domains.
- Assisting in representation learning processes, leading to data representations that encode critical factors from both data and formal semantics.
- Guiding training processes that capture critical semantics of data, with a representation that well generalizes a non-local prior over input space.
- Assisting in the resulting generation processes with expressive representation interpretations for high level semantics.

In this paper, we address the above goals with a semantic-rich deep learning framework that learns representations from both data distribution and formal semantics. Specifically, we propose an ontology-based deep restricted Boltzmann machine (OB-DRBM) model, in which formal ontology is used to guide architecture design of deep restricted Boltzmann machines (DRBM) [12], as well as to assist in their training and validation processes.

An ontology provides a formal representation of domain knowledge, through concepts, relationships, axiomatic constraints, and individuals. Figure 1 shows a sample ontology for news reports, recreational sports domain, used in one of our experiments. It contains a set of concepts for recreational sports in news reports, and relations between the concepts. Using a domain ontology, we can design an OB-DRBM model to learn a set of representations, each of which corresponds to a concept in the ontology. This set of representations learns to encode regularities from data with various semantic granularities for the current domain. For instance, using the news report ontology, we can learn representations that correspond to

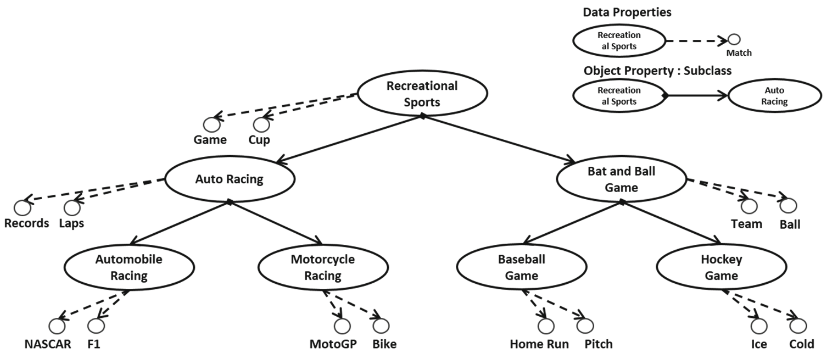


Fig. 1. A sample ontology for news reports, recreational sports domain. Each concept represents a type or category of recreational sport.

concepts, "*recreational sports*," "*auto racing*," and "*automobile racing*." Furthermore, our model provides a solution to semantic rich representation learning, in that representations learned for higher level semantics can support representation learning processes for their lower level subclass semantics. For instance, as shown in Fig. 2, in our model, representation learned for "*recreational sports*" can serve as a priori for the representation learning of "*auto racing*" and "*bat and ball games*." The inspiration for our OB-DRBM design primarily comes from the robustness theory of cognitive development process in biological neural networks.

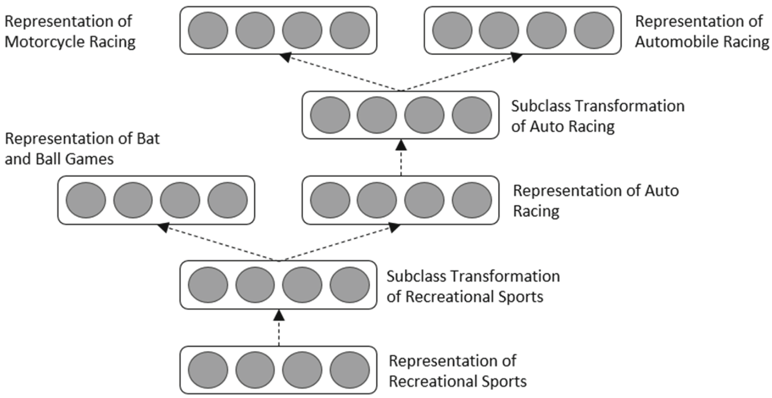


Fig. 2. Representation learning for news report ontology. For each concepts in this ontology, we use RBM layers to learn its data representation, which are further feed into RBM layers for related concepts, as auxiliary information.

In biological neural networks, many activities related to cognitive development process, such recognition and categorization, are often learned as a representation of a shared ontology [13]. Humans learn to categorize objects starting from early age by a hierarchal representation of object taxonomy in the world. The cognitive development process, for human being, usually begins with learning basic categories, such as ball, then progressively evolves into categories with more details, such as basketball [14]. Based on past knowledge and experience, a biological neural network learns by taking advantage of knowledge coming from previously learned categories, rather than learning from scratch. It leads to an efficient learning system that requires fewer samples to develop new generalization or ability promotion.

On the contrary, current deep neural networks not only require a large amount of data to make efficient learning and generalizations, but they also generalize poorly to data instances in a new but related domain [8,9]. Current deep neural networks not only require a large amount of data to make efficient learning and generalizations, but they also generalize poorly to data instances in a new but related domain [8,9]. Following the inspirations from robustness of human cognitive development process [14], we model representation learning in our model with a shared higher-level representation. We model representation learning in our model with a shared higher-level representation. We expect that representations learned for concepts at a higher semantic level, such as car and computer, can well assist in the process of learning representations, such as sedan and laptop, at more detailed semantic levels. It also renders our model the potential to explore the semantic relations between data instances, as well as the capability to learn a set of semantic rich representations with various semantic granularities.

Our contributions of this paper are as follows:

- We introduce a semantic-rich deep learning model, OB-DRBM, in which formal ontology has assisted in all stages of the deep learning process, including architecture design, training, and validation. Such architecture can learn a set of semantic-rich data representations from both data distribution and formal semantics. Representations learned correspond to concepts in a domain ontology, at various semantic levels.
- We propose corresponding training and validation methods, with assistance of inference and consistency-checking capabilities from ontologies and semantic reasoners. We show that our model leads to an improved performance, when compared with conventional deep learning models in text document classification tasks.

The remainder of this paper paper is structured as follows: Sect. 2 describes relevant previous works; Sect. 3 formally describes the architecture formulation of our model; in Sect. 4, we present our experiment result, when we apply OB-DRBM model to problems in various domains; in Sect. 5, we conclude our work by discussing potential future directions and their applications.

2 Related Work

In this paper, we propose to use formal ontology to assist in the deep learning process. Our OB-DRBM model learns a set of semantically related representations for each concept in a domain ontology. This set of representations also constitutes a formal semantics embedding based on both formal semantics and data distribution. Fields closely related to our model include, but are not restricted to, deep learning, knowledge engineering, and knowledge base embedding.

2.1 Deep Learning

In recent years, the rich set of deep neural network variations has lead to successes in numerous applications. Popular architectures of deep neural networks include, restricted Boltzmann machine (RBM) [12], convolutional neural networks (CNN) [1], and recurrent neural networks (RNN) [15]. RBM models have demonstrated exceptional performances for tasks with both labeled and unlabeled data [12]. CNN can effectively train data with topological structures and strong local correlations, such as image and speech [1]. RNN has been successfully applied on time series data and natural languages as a memory and latency model [15].

2.2 Knowledge Engineering

Knowledge engineering (KE) [16] is a research field that dedicates to develop techniques to build and reuse formal knowledge in a systematic way. In the past few decades, the proliferation of knowledge engineering has remarkably enriched the family of formal semantic representations. Ontology is one of the successful knowledge engineering advances. The encoded formal semantics in ontologies is primarily used for effectively sharing and reusing of knowledge and data. Prominent examples of domain ontologies include the Gene Ontology (GO [17]), Unified Medical Language System (UMLS [18]), and more than 300 ontologies in the National Center for Biomedical Ontology (NCBO [19]).

2.3 Knowledge Base Embedding

Recent research has developed methods to learn embeddings of knowledge base (KB) systems, such as WordNet, FreeBase, and DBPedia [20,21]. Entities in knowledge bases are embedded as low-dimensional vector representation. Syntactics, operations, and relations between entities are embedded as linear and bi-linear translations, matrix and matrix factorizations, and tensors. Bordes et al. [20] propose to learn vector-matrix embedding of knowledge base, in which knowledge bases are considered as graph models. Socher et al. [21] developed knowledge base embedding systems based on neural tensor networks for knowledge base completion. The key difference between our OB-DRBM model and previous knowledge base embedding model is, our model can learn embeddings from both data distributions and formal semantics, while previous methods learn only from a knowledge base.

3 Ontology-Based Deep Restricted Boltzmann Machine

In this section, we introduce our method to build an OB-DRBM model. We begin with a review of related techniques, including ontology in Sect. 3.1, semantic reasoner in Sect. 3.2, and restricted Boltzmann machine (RBM) in Sect. 3.3. We discuss the architecture design of our OB-DRBM model in Sect. 3.4 and corresponding training and validation methods in Sect. 3.5.

3.1 Ontology

Ontology [22] is an explicit specification of a shared conceptualization. The formal specification of an ontology can be defined as a quintuple, $\mathcal{O} = (C, P, I, V, A)$ where C, P, I, V, A are the set of classes, properties, individuals, property values and other axioms respectively [23]. Classes C, also referred to as concepts, describe the collections, concepts, types of objects and entities in a domain discourse. Properties P, also referred to as object properties, define relations between classes. Individuals I, are the instances or ground level objects of classes. Property values V, also referred to as data type properties, define features, attributes, parameter values that classes can have. Axioms A, define the ground truth of the domain discourse. The architecture design of our OB-DRBM model primarily uses the set of classes C and properties P in a domain ontology following the subclass relations in P. For concepts $c,\ s \in C$, we use $subclass(c, s)$, $superclass(c,\ s) \in P$ to denote the subclass and superclass relations between c and s. For each $c \in C$, $\pi(c) = \{s\ |superclass(s,\ c),\ s,\ c \in C\}$ and $\rho(c) = \{s\ |subclass(s,\ c),\ s,\ c \in C\}$ are used to denote the set of its subclass and superclass concepts.

3.2 Semantic Reasoner

A semantic reasoner [24] (also referred to as inference engine or reasoning engine) is a piece of software that infers logical consequences from a set of explicitly asserted facts or axioms. Prominent semantic reasoners of ontologies includes Pellet [25] and HerMit [26], and many more. It typically provides automated support for reasoning tasks such as deducting new knowledge, checking consistencies, verifying facts, and answering queries. Specifically, given a domain ontology $\mathcal{O}$ and a semantic reasoner $\mathcal{R}$, semantic reasoner can deduct an answer of query q based on the ontology $\mathcal{O}$ and axiom a, that $q = \mathcal{R}(O, a)$.

In our OB-DRBM model, the semantic reasoner is used in as a component for data semantics promotion and result validation. For a data instance $\{x, y\}$, a semantic reasoner $\mathcal{R}$ can return with promoted data instances with labels at a higher semantic level using $x \rightarrow \pi(y) = \mathcal{R}(\mathcal{O}, x \rightarrow y)$. For instance, for data instance $\{x, AutomobileRacing\}$, a semantic reasoner $\mathcal{R}$ can deduct with the valid promoted data instance, $\{x, AutoRacing\}$, using:

$$\frac{\begin{array}{c}\forall\ x\ AutomobileRacing(x)\\ \forall\ x\ AutomobileRacing(x) \rightarrow AutoRacing(x)\end{array}}{\forall\ x\ AutoRacing(x)}$$

By recursively applying $x \to \pi(y) = \mathcal{R}(\mathcal{O}, x \to y)$ k times, it can deduct with promoted data at even higher semantic levels, $x \to \pi^{(k)}(y) = \mathcal{R}(\mathcal{O}, x \to y)$.

Semantic reasoner can also validate the consistency of a set of axioms. For model with multiple representations and outputs, such as our OB-DRBM model; inconsistency can happen without consistency regulations from formal semantics. For instance, for classification outputs, $o_1 = x \to MotorcycleRacing$ and $o_2 = x \to BatAndBallGames$, a semantic reasoner can deduct with inconsistency state, $\perp = \mathcal{R}(\mathcal{O}, \{o_1, o_2\})$ using:

$$\frac{\dfrac{\begin{array}{c}\forall\ x\ MotorcycleRacing(x)\\ \forall\ x\ MotorcycleRacing(x) \to AutoRacing(x)\end{array}}{\forall\ x\ AutoRacing(x)} \qquad \forall\ x\ AutoRacing(x) \to \neg BatAndBallGames(x) \wedge BatAndBallGames(x)}{\perp.}$$

3.3 Deep Restricted Boltzmann Machine

A deep restricted Boltzmann machine (DRBM) is a deep neural network model with a stacking of many restricted Boltzmann machines (RBM) layers. RBM is a deep learning structure with bidirectionally connected binary stochastic processing units. Typically, a RBM contains a layer of visible units v and a layer of hidden units u, which are connected as a bipartite graph. RBM is a probabilistic graphic model that is based on an energy function defined on the exponential family. The joint probability that RBM assigned to visible units v and hidden units u are:

$$p(v, h) = \frac{exp(-E(v, h))}{Z}, \tag{1}$$

where $E(v, h)$ is a energy function defined on all RBM units, which indicates the degree of harmony of the network, Z is the partition function,

$$Z = \sum_{u,v} exp(-E(v, h)). \tag{2}$$

For RBM with binary visible units, $E(v, h)$ is defined as:

$$E(v, h) = -\sum_{i} a_i v_i - \sum_{j} b_j h_j - \sum_{i,\ j} v_i h_j w_{ij}. \tag{3}$$

For RBM with Gaussian visible units, $E(v, h)$ is defined as:

$$E(v,\ h) = -\sum_{i} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j} b_j h_j - \sum_{i,\ j} \frac{v_i}{\sigma_i} h_j w_{ij}. \tag{4}$$

where σ_i is the standard deviation for the Gaussian noise for visible unit i, a_i, b_i are the bias parameters for visible and hidden units and w_{ij} is the weight parameter of a RBM respectively.

Algorithm 1. OB-DRBM Architecture Design

Input: Ontology $\mathcal{O} = \{C, P, I, V, A\}$, Semantic Reasoner $\mathcal{R}$
Output: OB-DRBM structure $\mathcal{T}$

1: Let $r \in C$ be root concept of $\mathcal{O}$
2: Let s_c be an empty set
3: Add r into set s_c
4: **while** s_c is not *empty* **do**
5:     **for each** concept c in s_c **do**
6:         Initialize DRBM $\mathcal{D}_c$ for concept c
7:         Let $\rho(c) = \{s \mid s = \mathcal{R}(\mathcal{O}, subclass\ of\ c)\}$
8:         **if** $\rho(c)$ is not *empty* **then**
9:             Initialize $\mathcal{M}_c = mhmv\_layer(c, \rho(c))$
10:             Let $o_c = \{c \mid c \in C, c \notin \rho(c)\}$
11:             Let $t = \rho(c) \cup o_c$
12:             Initialize $\mathcal{S}_c = softmax\_layer(c, t)$
13:             Connect $\mathcal{S}_c$, $\mathcal{D}_c$ with $\mathcal{M}_c$
14:             Add $\mathcal{M}_c$, $\mathcal{S}_c$, $\mathcal{D}_c$ into $\mathcal{T}$
15:         **end if**
16:         Let $\pi(c) = \{s \mid s = \mathcal{R}(\mathcal{O}, superclass\ of\ c)\}$
17:         **if** $\pi(c)$ is not *empty* **then**
18:             Connect $\mathcal{D}_{\pi(c)}$ and $\mathcal{M}_{\pi(c)}$
19:         **end if**
20:         Let $s_c = \rho(c)$
21:     **end for**
22: **end while**
23: **return** $\mathcal{T}$

3.4 OB-DRBM Architecture Design

In this section, we present the architecture design of our OB-DRBM model. In Algorithm 1, we present the method of the model construction. Given an ontology $\mathcal{O}$ and a semantic reasoner $\mathcal{R}$, we compose the OB-DRBM model $\mathcal{T}$ following the subclass relations $\rho(c) \in P$ for each concept $c \in C$, for $C,\ P \in \mathcal{O}$. In Fig. 3, we show a sample OB-DRBM model following the sample news reports ontology in Fig. 1. The architecture design follows a top down process from higher level concepts to lower level concepts in C. The model construction process starts by adding the top level class $r \in C$ in the subclass hierarchy into the building sequence set s_c. For each concept $c \in s_c$, we first build a DRBM module $\mathcal{D}_c$ for the representation learning of concept c (lines 1–6). For top class r of the ontology, the DRBM module $\mathcal{D}_r$ takes only its own features as input. For other classes $c \in C, c \neq r$, the DRBM module $\mathcal{D}_c$ takes inputs from both its own features and transformed representations from its superclass modules $\mathcal{D}_{\pi(c)}$.

For each concept c and its corresponding DRBM module $\mathcal{D}_c$, we attach a semantic softmax layer $\mathcal{S}_c$, for semantic rich representation learning. The semantic softmax layer $\mathcal{S}_c$ is a layer that contains target output units at the corresponding semantic level. For each concept c, let $\rho(c)$ be the set of subclass concepts

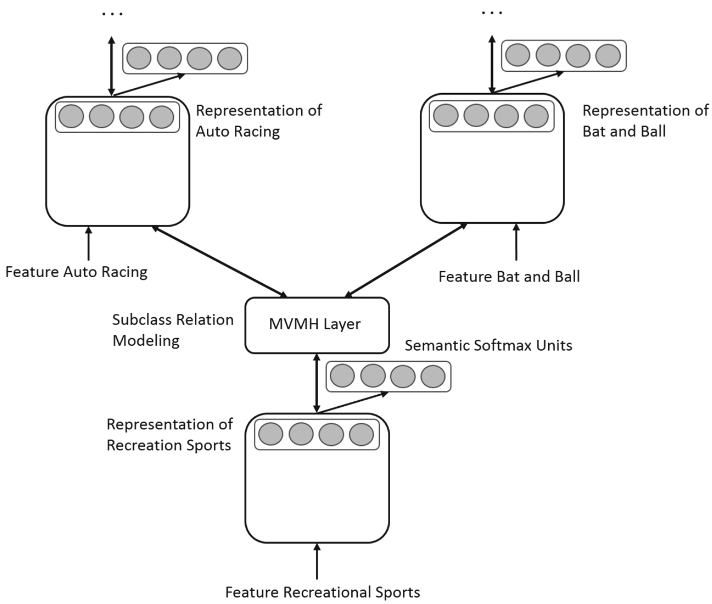


Fig. 3. An OB-DRBM architecture from news report ontology

of c. Each output unit in $\mathcal{S}_c$ corresponds to one concept in the subclass concept set, $\rho(c)$, plus one out of domain unit o_c. The unit o_c is used to model data that falls out of the domain of class c. For example, the semantic softmax layer $\mathcal{S}_{AutoRacing}$ contains three output units, for *AutomobileRacing*, *MotorcycleRacing*, *OutofDomain* respectively. For data instance $\{x,\ MotorcycleRacing\}$, the target output for $\mathcal{S}_{AutoRacing}$ is (0, 1, 0). At the training phase, through semantic reasoner query $\mathcal{R}(\mathcal{O}, x \rightarrow y)$, we can convert each labeled data $l = \{x,\ y\}$ to a set of promoted data instances, $l^{(k)} = \{x, \pi^{(k)}(y)\}$, for each semantic softmax layer.

For each concept c and its corresponding DRBM module $\mathcal{D}_c$, we also attach a multiple hidden multiple visible restricted Boltzmann machine (MHMV-RBM) layer $\mathcal{M}_c$ for subclass relation modeling. As shown in Fig. 4, a MVMH-RBM layer is a RBM variation designed to model the subclass transformation from a superclass to its subclasses. In our OB-DRBM model, each DRBM module $\mathcal{D}_c$ for concept c is attached to its own semantic softmax layer $\mathcal{S}_c$. The representation learned in $\mathcal{D}_c$ encodes the high level feature abstractions for concept c. Before feeding such a representation to subclass modules $\mathcal{D}_{\rho(c)}$, the MVMH-RBM layer learns a generative representation for both the input of subclasses features and representation in $\mathcal{D}_c$. The subclass representation and raw input are further feed into subclass modules $\mathcal{D}_{\rho(c)}$ as input.

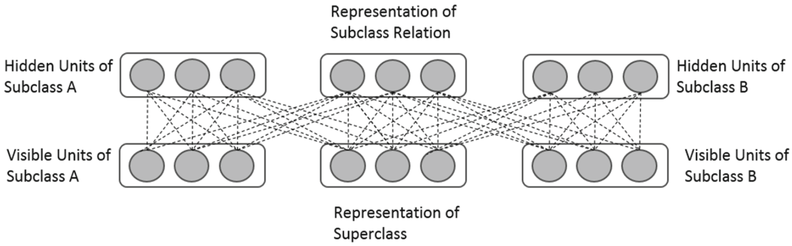


Fig. 4. A MHMV-RBM layer for subclass relation modeling

3.5 Training and Validation

We train our OB-DRBM model using a similar way as the conventional DRBM model. An OB-DRBM model was first trained with greedy module wised and layer wised contrastive divergence (CD) [27]. Then we use stochastic gradient descent across all semantic softmax output to further fine-tune our model with labeled data. In this process, we minimize the sum of cross entropy error for all softmax outputs of each concept in our model.

At the validation phase, the output of our OB-DRBM model contains a set of consistent outputs from all semantic softmax layer units. For example, $\hat{y} = \{$*Motorcycle Racing, Auto Racing, Recreational Sports*$\}$ is a consistent output for input data, $\{x,\ MotorcycleRacing\}$. We enforce this consistency using a logistic regression across all semantic softmax output configurations with consistency validation from a semantic reasoner. Specifically, let $\mathcal{S}$ be the set of all softmax output values, the set of outputs is,

$$\hat{y} = \underset{s \subset \mathcal{S}}{\operatorname{argmax}} \frac{\prod_{c \in s} f_c(x, w)[\mathcal{R}(\mathcal{O}, x \to s)]}{\sum_{s \subset \mathcal{S}} \prod_{c \in s} f_c(x, w)[\mathcal{R}(\mathcal{O}, x \to s)]}, \tag{5}$$

in which $f_c(x, w)$ is the softmax confidence value for unit c, $[\mathcal{R}(\mathcal{O}, x \to s)]$ is the activation function that ensures a valid output configuration.

4 Experiment

We present experiments on two problems related to text documents: topic classification and sentiment analysis. For selected text documents, we adopt a continuous bag of word model [28] in our experiment to convert text documents into continuous vector representations. From the frequent word set, we remove stop words and the most frequent 100 words, then keep the next 5000 most frequent words. In our experiments, we adopt the bag of word model primarily for its simplicity. We understand that the bag of word model might not be the best fit and state of the art approach for the datasets to which we have applied our method on. However, our primary goal is to explore the effect of formal semantics in deep learning process. We verify our theory by comparing our OB-DRBM model with conventional DRBM model under the same context, including data distribution, meta parameters, training time and algorithms, and so on. In all experiments, we divide the dataset into 70 % training, 15 % validation, and 15 % testing. The number of iterations over the training set was determined using early stopping according to the validation set classification error with an additional 100 iterations.

4.1 News Topic Classification

We first evaluated our model on the news topic classification problem on 20 Newsgroups dataset [29]. The data is organized into 20 different newsgroups, each corresponding to a different topic, across four domains of *computer company*, *recreational sports*, *science*, and *public talks*. We define domain ontologies for each of those domains, based on the natural taxonomy relations of the topics. We have shown one example domain ontology defined for this dataset in the recreational sports domain in Sect. 1, Fig. 1. Other domain ontologies defined for our experiments can be found in our website [30].

Table 1 gives the classification performances on the four topic domains. Our OB-DRBM model outperforms the conventional DRBM models in 3 out of the 4 domains, including *company*, *sports*, and *social*, In the *science* domain, DRBM model outperforms our model but only by a less than 1 % margin. This is mostly because the 4 topics in the *science* domain, *sci.electronics*, *sci.medicine*, *sci.space*, and *sci.crypt* share very few common characteristics. The best domain ontology that fits with the data is an ontology with a flat structure. In this case, our OB-DRBM model cannot benefit from the shared representation of superclass in this ontology.

Table 1. Classification performance on news topics

Topic domain	OB-DRBM	DRBM
Company	**77.46** %	75.83 %
Sports	**82.11** %	79.57 %
Social	**74.20** %	72.69 %
Science	70.46 %	**71.32** %

4.2 Sentiment Analysis Datasets

We further conduct our experiment upon document datasets on sentiment analysis tasks. We test our OB-DRBM model on the Pang/Lee movie review data [31] and sentiment analysis dataset from sentiment tree bank [32]. In both datasets, movie reviews are labeled as four categories, *positive*, *neutral positive*, *neutral negative*, and *negative*. Table 2 gives the classification performances on sentiment analysis tasks. In both datasets, our OB-DRBM model outperforms conventional DRBM model by a large margin.

Table 2. Classification performance on sentiment analysis

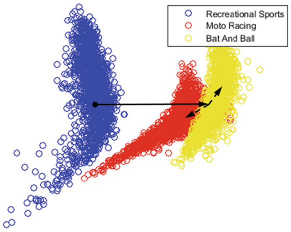
Data set	OB-DRBM	DRBM
Pang/Lee	**68.09** %	64.20 %
Sentiment tree bank	**60.19** %	54.45 %

4.3 Data Simulation of Formal Semantics Embedding

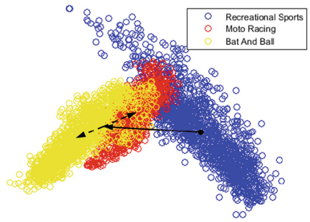
One primary motivation of our work is to learn a structured set of representations from both the formal semantics and the data distribution. We expect this set of semantic rich representations can encode regularities of the data at various semantic levels, such that representations of higher-level semantics can encode the common data regularities of their lower-level subclass semantics. We exam our hypothesis through visualization of the representation the learning process. In Fig. 5, we present the low-dimensional principle component analysis (PCA) embedding of the representations learned in our OB-DRBM model at various epochs of the supervised-training process. It shows the representation embeddings of three concepts with subclass relations, "*recreational sports*, "*automobile racing*," and "*bat and ball games*" in our recreational sports ontology. Before the supervised-training process, the OB-DRBM model was first trained with unsupervised-training using contrastive divergence (CD) [27].

In Fig. 5(a), we show the set of representations learned in our OB-DRBM model after the unsupervised-training. At this phase, the model can only learn from the data distribution. There is neither any data semantics, nor any formal knowledge semantics involved during this phase. After the unsupervised-training, representation learned for superclass and subclasses are roughly of the same distribution. Without the direction of formal semantics, each of the three representations plays a similar role in the model. At the 500th epoch, as shown in Fig. 5(b), the distributions of the three data representations are still similar. However, with assistance of formal semantics and labeled data, the representation of superclass, "*recreational sports*," as diverged into a different principle components compared with the representations of its subclasses.

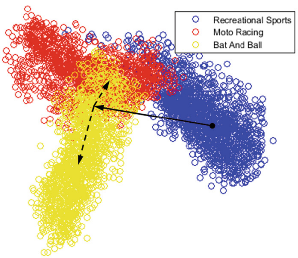
At the 3000th epoch, as shown in Fig. 5(c), principle components of the representations for the two subclass concepts, "*moto racing*" and "*bat and ball*," start to show difference as well. Distinction of distributions has started to emerge between the representation of the superclass "*recreational sports*" and the two subclasses. At the 5000th epoch, as shown in Fig. 5(d), the model learns a set of data representations with three distinct principle components and distributions. At this stage, the representations of the superclass and the subclasses has encoded data representations with different levels of semantics. We can see through this process, how the set of semantic rich data representations influence each other through the assistance of formal semantics. When the superclass representation starts to model the common semantics of "*recreational sports*" gradually, the representations of the two subclasses were set free to learn its local semantics as well.



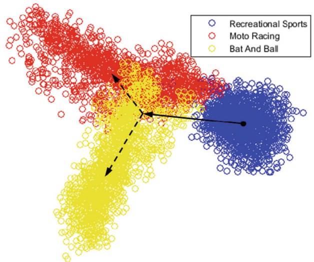
(a) Embedding at the 10th epoch



(b) Embedding at the 500th epoch



(c) Embedding at the 3000th epoch



(d) Embedding at the 5000th epoch

Fig. 5. Visualization of OB-DRBM representations.

5 Conclusion and Future Work

We have evaluated the potential of semantic rich deep learning using our OB-DRBM model. We have demonstrated that, with assistance of formal semantics, deep learning models can learn a set of semantic rich representations from both formal semantics and data. This set of representations constitute a structured embedding of formal knowledge under the data distribution. They also lead to improved performances in document classification tasks.

For future work, we would like to investigate the embedding learning of formal semantics in more forms, such as convolutional neural networks, or matrix vector embeddings. We would like to explore the potential to learn from unsupervised-data with assistance of formal semantics, as well.

Acknowledgment. This work is supported by the NIH grant R01GM103309. We acknowledge Ellen Klowden for her contributions to the manuscript. We also thank anonymous reviewers for their constructive comments, which helped improve the paper.

References

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
2. Le, H.S., Oparin, I., Allauzen, A., Gauvain, J., Yvon, F.: Structured output layer neural network language models for speech recognition. IEEE Trans. Audio Speech Lang. Process. **21**(1), 197–206 (2013)
3. Mikolov, T., Deoras, A., Kombrink, S., Burget, L., Cernockỳ, J.: Empirical evaluation and combination of advanced language modeling techniques. In: Annual Conference of the International Speech Communication Association, pp. 605–608 (2011)
4. Srivastava, N., Salakhutdinov, R.R.: Multimodal learning with deep Boltzmann machines. In: Advances in Neural Information Processing Systems, pp. 2222–2230 (2012)
5. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)
6. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Proceedings of the 13th European Conference on Computer Vision, pp. 818–833 (2014)
7. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
8. Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., Fergus, R.: Intriguing properties of neural networks (2013). arXiv preprint arXiv:1312.6199
9. Nguyen, A., Yosinski, J., Clune, J.: Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 427–436 (2015)
10. Dou, D., Wang, H., Liu, H.: Semantic data mining: a survey of ontology-based approaches. In: IEEE International Conference on Semantic Computing, pp. 244–251 (2015)
11. Balcan, N., Blum, A., Mansour, Y.: Exploiting ontology structures and unlabeled data for learning. In: Proceedings of the 30th International Conference on Machine Learning, pp. 1112–1120 (2013)
12. Salakhutdinov, R., Hinton, G.E.: Deep Boltzmann machines. In: International Conference on Artificial Intelligence and Statistics, pp. 448–455 (2009)
13. Kolb, B., Whishaw, I.Q.: Fundamentals of Human Neuropsychology. Macmillan, London (2009)
14. Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., Boyes-Braem, P.: Basic objects in natural categories. Cogn. Psychol. **8**(3), 382–439 (1976)
15. Socher, R., Lin, C.C., Manning, C., Ng, A.Y.: Parsing natural scenes and natural language with recursive neural networks. In: Proceedings of the 28th International Conference on Machine Learning, pp. 129–136 (2011)
16. Russell, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach, 2nd edn. Pearson Education, Upper Saddle River (2003)
17. Consortium, T.G.O.: Creating the gene ontology resource: design and implementation. Genome Res. **11**(8), 1425–1433 (2001)
18. Lindberg, D., Humphries, B., McCray, A.: The unified medical language system. Methods Inf. Med. **32**(4), 281–291 (1993)

19. NCBO: The National Center for Biomedical Ontology. http://www.bioontology.org/
20. Bordes, A., Weston, J., Collobert, R., Bengio, Y.: Learning structured embeddings of knowledge bases. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 301–306 (2011)
21. Socher, R., Chen, D., Manning, C.D., Ng, A.: Reasoning with neural tensor networks for knowledge base completion. In: Advances in Neural Information Processing Systems, pp. 926–934 (2013)
22. Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquis. **5**(2), 199–220 (1993)
23. Wimalasuriya, D.C., Dou, D.: Components for information extraction: ontology-based information extractors and generic platforms. In: Proceedings of the 19th ACM Conference on Information and Knowledge Management, pp. 9–18 (2010)
24. Dentler, K., Cornet, R., Ten Teije, A., De Keizer, N.: Comparison of reasoners for large ontologies in the OWL 2 EL profile. Seman. Web **2**(2), 71–87 (2011)
25. Sirin, E., Parsia, B., Grau, B.C., Kalyanpur, A., Katz, Y.: Pellet: a practical OWL-DL reasoner. Web Seman. **5**(2), 51–53 (2007)
26. Motik, B., Shearer, R., Horrocks, I.: Hypertableau reasoning for description logics. J. Artifi. Intell. Res. **36**, 165–228 (2009)
27. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313**(5786), 504–507 (2006)
28. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space (2013). arXiv preprint arXiv:1301.3781
29. Lang, K.: Newsweeder: learning to filter netnews. In: Proceedings of the International Conference on Machine Learning, pp. 331–339 (1995)
30. AIMLAB: Ontologies. http://aimlab-server.cs.uoregon.edu/ontologies
31. Pang, B., Lee, L.: Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the Annual Meeting on Association for Computational Linguistics, pp. 115–124 (2005)
32. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, p. 1642 (2013)