

# Towards Unsupervised Knowledge Extraction

Dorothea Tsatsou<sup>a,b</sup>, Konstantinos Karageorgos<sup>a</sup>, Anastasios Dimou<sup>a</sup>, Javier Carbo<sup>b</sup>, Jose M. Molina<sup>b</sup> and Petros Daras<sup>a</sup>

<sup>a</sup>Information Technologies Institute (ITI), Centre for Research and Technology Hellas (CERTH), 6th km Charilaou-Thermi Road, 57001, Thessaloniki, Greece

<sup>b</sup>Computer Science Department, University Carlos III of Madrid, Av. Universidad 30, Leganes, Madrid 28911, Spain

## Abstract

Integration of symbolic and sub-symbolic approaches is rapidly emerging as an Artificial Intelligence (AI) paradigm. This paper presents a proof-of-concept approach towards an unsupervised learning method, based on Restricted Boltzmann Machines (RBMs), for extracting semantic associations among prominent entities within data. Validation of the approach is performed in two datasets that connect language and vision, namely Visual Genome and GQA. A methodology to formally structure the extracted knowledge for subsequent use through reasoning engines is also offered.

## Keywords

knowledge extraction, unsupervised learning, spectral analysis, formal knowledge representation, symbolic AI, sub-symbolic AI, neuro-symbolic integration

## 1. Introduction

Nowadays, artificial intelligence (AI) is linked mostly to machine learning (ML) solutions, enabling machines to learn from data and subsequently make predictions based on unidentified patterns in data, taking advantage of neural network (NN)-based methods. However, AI is still far from encompassing human-like cognitive capacities, which include not only learning but also understanding<sup>1</sup>, abstracting, planning, representing knowledge and logically reasoning over it. On the other hand, Knowledge Representation and Reasoning (KRR) techniques allow machines to reason about structured knowledge, in order to perform human-like complex problem solving and decision-making.

AI foundations propose that all the aforementioned cognitive processes (learning, abstracting, representing, reasoning) need to be integrated under a unified strategy, in order to advance to

---

In A. Martin, K. Hinkelmann, H.-G. Fill, A. Gerber, D. Lenat, R. Stolle, F. van Harmelen (Eds.), *Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering (AAAI-MAKE 2021)* - Stanford University, Palo Alto, California, USA, March 22-24, 2021.

✉ dorothea@iti.gr (D. Tsatsou); konstantinkarage@iti.gr (K. Karageorgos); dimou@iti.gr (A. Dimou);

jcarbo@inf.uc3m.es (J. Carbo); molina@ia.uc3m.es (J.M. Molina); daras@iti.gr (P. Daras)


ORCID 0000-0003-0554-9679 (D. Tsatsou); 0000-0002-5426-447X (K. Karageorgos); 0000-0003-2763-4217 (A. Dimou);

0000-0001-7794-3398 (J. Carbo); 0000-0002-7484-7357 (J.M. Molina); 0000-0003-3814-6710 (P. Daras)

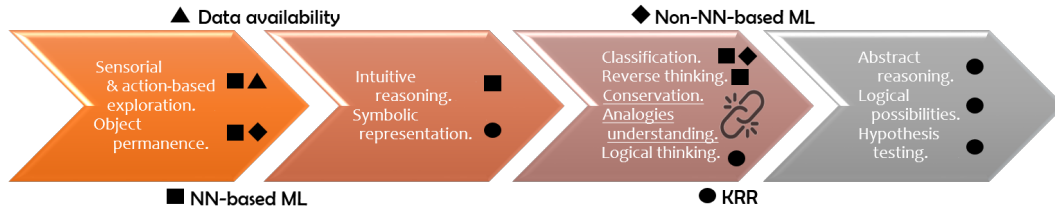


© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

<sup>1</sup>"Understanding" is a term that touches complex human cognitive capacities and is equivocal when applied to machine intelligence. In the premises of this paper, we use the term to denote 'digestion' of disparate information to capture the major premises (semantics) of an entire domain, or of particular content, tasks or any other contextual premise.



**Figure 1:** Analogies of cognitive development theory (Piagetian-inspired) to modern AI.

stronger AI. This integration involves the progress from intuitive, sub-symbolic intelligence to symbolic, logic-based cognitive processes. Bridging sub-symbolic (connectionist) AI with symbolic AI has long been considered crucial towards effective AI solutions [1].

Recognizing the need of symbolic and sub-symbolic integration, the work of this paper is inspired by the amalgamation of different learning and epistemology theories on human cognitive development. Besides their functional differences, most theories converge to a developmental process by which human cognition evolves. In this process, sub-symbolic learning is fundamental (both in the sense of rudimentary, as well as in the sense of necessary) to obtain symbolic function, subsequently used for concrete or abstract logic and reasoning [2].

The current success of NN-based sub-symbolic learning and the plethora of symbolic KRR solutions long available, signify that AI is now equipped to complete its development cycle, but with one missing link between two ends: non-manual acquisition of symbolic knowledge based on the distillation of information hidden in sub-symbolic models. Figure 1 portrays the analogies with the Piagetian [3]<sup>2</sup> cognitive development stages to AI’s processes.

The work of this paper corresponds to analogies understanding, i.e. making connections between concepts. Consequently, we propose an unsupervised approach for extracting knowledge based on the patterns formed in trained neural networks, namely from Restricted Boltzmann Machines (RBMs) [4]. The extracted knowledge can be represented formally and thus shared, used and re-used for logical inference over any domain. The idea is in its early fruition stage and the paper performs a sanity check over the proposed approach, with an interest in its applicability to different domains and/or data, and presents a concrete plan for the evolution of the approach.

To this end, Section 2 offers an overview of related work; Section 3 presents the implemented method, along with first experimental results and observations; Section 4 provides a conclusion over the initial approach and a concrete overview of future work.

## 2. Related Work

Integration of sub-symbolic and symbolic methods pertains to three major research directions: a) symbolic representation learning through neural approaches; b) induction of structured

<sup>2</sup>Piaget’s theory has received criticism in terms of terminology, stage transition and contents, causation of gained attributes, etc. This is the reason why the specific stage names are not used in Figure 1. However, it does provide a concrete flow of the development of cognitive capacities in humans, useful to depict the analogies in human vs machine intelligence developmental processes.

knowledge and/or some logical inference capacities in neural approaches, e.g. [5], [6]; c) hybrid methods that combine symbolic and sub-symbolic solutions to solve different parts of specific problems, e.g. [7], [8].

The scope of this paper lies in (a) - symbolic representation learning. This can be further divided to: i) recognition of pertinent symbolic representations of salient concepts within a domain and/or their hierarchy; ii) recognition of prominent relations that associate particular concepts; iii) pattern mining for extracting particular associative rules within a domain of discourse.

Representation learning focuses on identifying/automatically constructing the input features needed to perform a specific ML task. In computer vision, representations learned often do not have any symbolic manifestation, rather remain unstructured, black box vehicles in ML-based classification and feature detection [9], with few approaches aligning representations learned to specific symbolic labels [10].

In NLP, representation learning revolves mostly around the construction of word embeddings, associating words of high lexical proximity. Symbolic representation of natural language components is inherent, often accompanied with underlying semantics [11]. Several supervised methods have tried to move beyond mere term proximity identification to deeper semantics recognition, usually delving into encoding taxonomic relations between words within term embeddings [12].

Symbolic representation of non-taxonomic relations, however, is one of the most pertinent tasks towards structured knowledge acquisition. To this end, [13] employ modular neural networks, combining different NN models (e.g. image and text) with common features fused in a cumulative model, thus allowing to learn how textual relations associate objects in images.

The Neuro-Symbolic Concept Learner [14], not only learns object representations in scenes through natural supervision – i.e. no labeled data required – but also learns non-specified binary relations between recognized objects, combining visual and textual cues. Ultimately, knowledge learned is formalized and reasoned upon, in a question-answering task. The method is however yet confined to very few specific objects and predefined relations between them. Still, it paves the way towards pragmatic neural extraction of knowledge and subsequently towards symbolic inferencing over new knowledge.

The Neural Symbolic Cognitive Agent (NSCA) [15] uses a Recurrent Temporal Restricted Boltzmann Machine (RTRBM) [16] to learn complex temporal relations between data and subsequently formalize these relations into propositional rules, to be used for subsequent inferencing. The method has been applied to a restricted domain task and only to enrich existing knowledge with non-persistent, non-verified (to be exact, not needing verification) knowledge, however it unveils the capacity of RBMs as a powerful neural method for obtaining associations and rules among entities.

[17] also employ Restricted Boltzmann Machines (RBMs) in order to extract fuzzy rules, encapsulating the uncertainty/vagueness of probabilistic machine learning within logic-compliant rules, something missing from the crisp NSCA method. However, this method lacks comprehensive description of the formalization method for the extracted rules.

The hybrid Differential Inductive Logic Programming (DILP) method [18] combines differential neural-based learning with traditional inductive logic programming to learn and subsequently predict the less than/greater than relationships among visual data of the MNIST

digits database and formalizing them into logical rules, however still remaining a crisp and supervised approach.

### 3. Symbolic Relational Representations Extraction

Inspired by [15] and [17], the presented approach aims at extracting intricate knowledge, i.e. non-predefined associations among data, through RBMs [19] and symbolically representing them for further use in knowledge-based inferencing. RBMs were elected due to their capacity for unsupervised learning of probabilities over their input, which can in turn be further analysed to yield fuzzy associations among the input entities.

The proposed approach uses a state-of-the-art algorithm for RBMs, following the benchmarking implementation of RBMs for Deep Belief Nets (DBNs) of [19], but with a more efficient training algorithm than the first implementation, i.e. an extension of the original Contrastive Divergence (CD) algorithm, namely Persistent Contrastive Divergence (PCD) [20].

RBMs compute the probability distribution over pairs of visible and hidden vectors,  $V$  and  $H$  respectively, by the equation:

$$P(v, h) = \exp(v^\top b_v + h^\top b_h + v^\top W h) / Z, \text{ with } V \in \{0, 1\}^{N_V}, H \in \{0, 1\}^{N_H} \quad (1)$$

where  $b_v$  and  $b_h$  are biases (initialized uniformly in our implementation) of the visible and hidden vectors respectively,  $W$  is the matrix of connection weights and  $Z = b_V + b_H + W$  is a partition function that ensures the validity of the probability distribution [16]. The training algorithm (PCD) for the network finds the states of visible ( $V$ ) and hidden ( $H$ ) units that lower the total energy of the network ( $E$ ), thus maximizing the likelihood of correlation between visible units based on their connection through the network's hidden units, as described in [20].

$$E(v, h) = - \sum_{i,j} w_{ij} v_i h_j - \sum_i b_i^v v_i - \sum_j b_j^h h_j \quad (2)$$

Throughout training we monitored the energy gap between the training and validation sets to avoid overfitting, as per [4]. For each experiment the number of hidden units was set equal to the number of the visible ones (300 for Visual Genome and 250 for GQA, re Section 3.1.1), while the learning rate was set at  $10^{-4}$ .

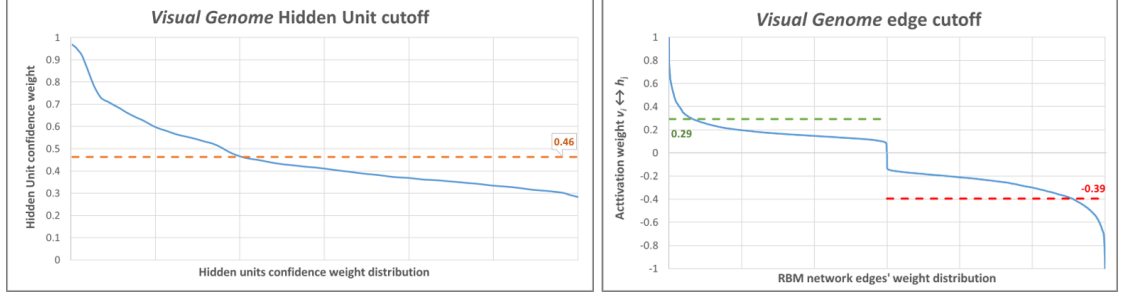
#### 3.1. RBM Training

The approach aims at examining the capacities of RBMs to extract persistent, common-sense, semantic relations among classes in datasets of multi-labelled annotated images. The reason why visually-oriented relations' extraction was opted for the first approach (as opposed to e.g. natural language) is the rudimentary nature and distinguishability of visual data interrelations (as opposed to the complexity and semantic ambiguity of natural language).

In order to examine the domain independence capacities of the approach, the experiment was performed in two distinct datasets.

##### 3.1.1. Experiment Setup

In order to train our network we used two popular structured image datasets, Visual Genome [21] and GQA [22]. Both contain rich annotations about the objects (object classes) present in an



**Figure 2:** Visual Genome thresholding. Left: Hidden units' confidence-based curve and cutoff point. Right: Activation weights (for maintained hidden units) distribution curve and cutoff points.

image, as well as their semantic relationships. As input, we pass a binary vector per sample (dataset image), denoting the existence of one or more object classes in that sample. Each image may contain multiple object classes, making the input an n-hot vector. Although object names (annotations) are free-form in both datasets, they both offer the means to map names into a consistent object class dictionary. Therefore, for Visual Genome we assigned each object name to the provided WordNet [23]-based synset it belongs to and converted each object name to a WordNet lemma for GQA.

Since both datasets contain a large number of object classes, we only used those with more than  $\approx 1000$  appearances, in order to restrict the computational cost of training, totaling to 300 classes for Visual Genome and 250 for GQA.

### 3.1.2. RBM Network Pruning

The reconstructed input yielded several relations among all  $v_i \in V$  under each  $h_j \in H$ . However, not all hidden units bore the same significance, since several pertained to connections with significantly low activation weights for all  $v_i \xleftrightarrow{w_{ij}} h_j$ . To examine their significance, each  $h_j \in H$  was assigned a normalized confidence weight  $cw_{h_j}, j \in N_H$ , based on the method employed in [24], as seen in Eq. 3.

Based on [25], a cutoff mechanism was devised in order to maintain only the "beneficial"  $h_j$ , i.e. the hidden units whose relative computational cost would still benefit the analysis, while the rest were discarded as noise. To this end, the *elbow/knee* [25] pertained the *cutoff point for the hidden units* to be maintained (Eq. 4).

$$cw_{h_j} = \sum_i |w_{ij}| \quad (3) \quad K_f(cw_H) = \frac{f''(cw_H)}{(1 + f'(cw_H)^2)^{15}} \quad (4)$$

Similarly, a cutoff point was devised in order to maintain the "beneficial" visible units per each hidden unit  $\forall h_j : v_i \leftrightarrow h_j$ . To this end, the *elbow/knee*  $K_f(w_{v_i}), w_{v_i} \geq 0$  and  $K_f(w_{v_i}), w_{v_i} < 0$  was computed for all activation weights in the RBM, as the cutoff point to prune all visible units within each hidden unit that bore a low, non-beneficial, activation weight.

A graphical representation of the hidden unit confidence and activation distribution curves and their elbow-based cutoff points for Visual Genome can be seen in Figure 2.

### 3.2. Spectral Analysis of RBM Network

To explore the patterns formed in the RBMs' network structure, spectral analysis was employed. More precisely, a graph representation  $G_{RBM} = (V_{RBM}, E_{RBM})$  was created, per each of the datasets, based on the pruned RBM network's structure. The graphs pertain of weighted vertices, interconnected by weighted edges, where  $wv_{RBM_k} = cw_{h_j}, \forall v_{RBM_k} \in H$  and  $wv_{RBM_k} = \widetilde{freq}(v_i), \forall v_{RBM_k} \in V$ , where  $\widetilde{freq}(v_i) \in [0, 1]$  is the normalized frequency of  $v_i \in V$  in the training dataset. Regarding edges, weight  $we_{RBM} = w_{ij}v_i(h_j)$  is assigned.

In order to examine the most prominent associations arising among the interconnected vertices, taking also advantage of the vertex and edge weights, a combinatorial Laplacian matrix with vertex and edge weights [26] was used, with  $\mathcal{L} = BTB^*$ , where  $B$  is the weighted incidence matrix and  $T$  is the weighted edge incidence diagonal of each graph.

In order to explore the RBM results against the ground truth, the results of the RBM's spectral analysis were compared against the spectral analysis of the co-occurrence between object classes in the ground truth. To this end, a graph representation of the ground truth per each dataset was devised, as  $G_{GT} = (V_{GT}, E_{GT})$ <sup>3</sup>. The graphs again bear weighted vertices and edges, with  $wv_{GT_k} = freq(v_n), \forall v_{GT_k} \in V$ , where  $freq(v_n)$  is the frequency of  $v_n \in V$  in the training dataset. Edge weights are designated as  $we_{GT_k} = \sum_{n,m} cooc(v_n, v_m)$ , where  $cooc(v_n, v_m) \in \{0, 1\}^{N_v}$  denotes the cooccurrence between two classes in an image of the dataset.

Lastly, to extract the associations, spectral clustering based on  $\mathcal{L}$  was performed, using DBScan [27] with a *relative gap* [28], with  $minPts = 2$  as even pair-wise associations are relevant to our goal,  $eps = 0.75$  set empirically as the optimal radius and  $relGap = \|x\|_2$  of  $\mathcal{L}$  found to be the optimal relative gap for both RBM-based graphs, but also Ground Truth graphs. Subsequent work aims at also automating retrieval of the optimal  $eps$  based on  $\mathcal{L}$ 's spectral properties.

### 3.3. Results

The results of this analysis revealed the capacity of RBMs to extract prominent relations among visible units. The extracted clusters of interrelated classes can be seen in the Appendix, in sections A.2 and A.1.

#### 3.3.1. Observations

Spectral analysis of the ground truth data graphs already yields results that capture valid common-sense interrelations among input classes. However, spectral analysis over the RBMs has revealed, in many cases, different and more intricate relations than in the ground truth data. Ground Truth graphs are found (as expected) to solely depict visual co-occurrence. This is apparent by the difference in number of clusters, as well as from the non-correspondence in semantics that RBM-based spectral analysis produces as opposed to Ground Truth analysis.

Moreover, within RBM-produced clusters, abstractions *and* specializations that are globally (not only visually) applicable for the observed semantic senses within their domain have been revealed. For example (see Appendix A.2), RBM analysis in Visual Genome was able to discern

<sup>3</sup>  $V_{GT}$  denotes all classes in the dataset and thus coincide with  $V_{RBM}$ , effectively making  $V_{GT}, V_{RBM} \in V$



**Table 1**

Evaluation results of RBM-produced and Ground

	VG RBM	VG GT	GQA RBM	GQA GT
Number of clusters	28	14	12	11
Semantic validity per cluster	0.98	0.99	0.958	0.962
Clusters with shared/similar semantics		8		0
Precision in common semantics clusters	1.00	0.99	N/A	N/A
Recall in common semantics clusters	0.91	0.84	N/A	N/A

between things related to food per se and objects found in an eating area, while the relevant Ground Truth analysis bundled most of these objects under a single cluster related to food. Several abstract relations such as e.g. the ones between logos, letters and design, between words, writing and signs, among different body parts, facial parts, animal parts, vehicle parts, etc have only been captured through RBM analysis.

Most interestingly, other non-direct properties were revealed based on seed classes that served as common denominators over related objects. For example, a RBM-produced Visual Genome cluster pertains to lady, dress, skirt, bag, child, male child. Upon further inspection based on VG's semantic relations among data reveals that this is in fact a two-sense tree, with lady being the common denominator and dress, skirt, bag constituting a semantic branch regarding clothing related to women, while child, male child constituting a disjoint branch<sup>4</sup> of types of persons frequently related to women. Such information can enable the unsupervised retrieval of hierarchy or meronymy relations among classes.

### 3.3.2. Evaluation

Although it is very difficult to evaluate the validity of the produced interrelations due to the lack of a golden standard and to contextual subjectivity, the produced clusters of interrelated objects based on the RBM-graph spectral analysis were compared against the clusters produced from the ground truth-graph spectral analysis, in terms of semantic validity.

Refraining from self-justification of the results, two independent observers studied the produced clusters in order to identify the 'common sense' semantic contexts of the produced clusters in both RBM and ground truth graphs.

Moreover, the Visual Genome dataset's RBM-based spectral analysis produced some clusters of similar semantics with its ground truth, which gave the opportunity to measure precision and recall among the semantically similar clusters. The GQA results however did not yield any semantically similar clusters. The results can be seen in Table 1.

## 3.4. Representation of Extracted Knowledge

Since the semantics of the produced "relation clusters" are not disentangled at this stage, every member of each cluster is considered as generically related to each other member of the cluster.

Therefore, as a first step, a generic *isRelatedTo* object property can be used to construct generic symbolic rules that express the interrelations among cluster members.

<sup>4</sup>The members of each branch do not co-occur in the dataset.

To this end, the interrelations extracted can be expressed in Description Logics (DLs) notation as propositional axioms of the form:

$$C_{i_n} \sqsubseteq \forall(isRelatedTo.C_{i_m}) \quad (5)$$

where  $C$  is a class in the dataset,  $i$  is a single cluster produced through the spectral analysis of the RBM-based graph and  $n, m$  are classes  $\in$  cluster  $i$ .

In first order logic (FOL), this translates to:

$$\forall x.C_{i_n}(x) \rightarrow (\forall y.isRelatedTo(x, y) \rightarrow C_{i_m}(y)) \quad (6)$$

Based on this representation, several inference tasks can provide richer information while reasoning over given facts in a particular domain. For example, a visual question answering system may use the *isRelatedTo* property to query, based on a DL or FOL reasoner, additional aspects related to an image, given an instance  $x$  of a class  $C_{i_n}(x)$  retrieved or explicitly annotated in an image  $y$ , by instantiating accordingly *isRelatedTo*( $x, y$ ).

For example, given the VG RBM-produced cluster `word.n.01`, `writing.n.01`, `sign.n.02`, the following axioms may be produced:

$$word \sqsubseteq \forall(isRelatedTo.writing) \quad (7a) \quad sign \sqsubseteq \forall(isRelatedTo.word) \quad (7d)$$

$$writing \sqsubseteq \forall(isRelatedTo.word) \quad (7b) \quad sign \sqsubseteq \forall isRelatedTo.writing \quad (7e)$$

$$word \sqsubseteq \forall(isRelatedTo.sign) \quad (7c) \quad writing \sqsubseteq \forall(isRelatedTo.sign) \quad (7f)$$

Based on this knowledge, a question-answering system using a knowledge-based (deductive) inference engine can answer the question "What is related to this image?" given an image *imagename*, where a *sign*(*infosign*) was identified, by grounding the *isRelatedTo* property for *isRelatedTo*(*infosign*, *imagename*). Through axioms 7d and 7e, we get:

$$sign(infosign) \sqsubseteq \forall isRelatedTo(infosign, imagename.word(y)) \models word(imagename)$$

$$sign(infosign) \sqsubseteq \forall isRelatedTo(infosign, imagename.writing(y)) \models writing(imagename)$$

Effectively, the contents of the image has some relation to words and writing, given that sign exists in the image.

## 4. Conclusions and Future Work

This paper presented an early neuro-symbolic integration approach for unsupervised symbolic knowledge extraction from trained neural networks, by using RBMs to uncover persistent semantic associations among concepts found in multi-labelled images of the Visual Genome and GQA datasets. Valuable insights arose in the process about the capacity of RBMs and spectral analysis to uncover relational knowledge from data.

The purpose of this first proof-of-concept work is to determine the validity of the assumption that the obscure patterns that are formed in trained NNs may be captured symbolically, to some extent, as structured knowledge, able to construct, update or complement knowledge bases. The goal of this process is to be task-independent and that the knowledge acquired can be further re-used, shared and used for complex reasoning. The approach aims to be applicable to any domain for which data is available, leverage the need for huge data and yield consistent results in terms of extracted propositions' accuracy.

The next steps will delve into examining whether the same applies in deeper architectures and what more/else deeper networks may reveal, by expanding the approach from RBMs to



DBNs. The directed nature of the added layers in DBNs<sup>5</sup>, in combination with the capacity of the current approach to recognize the k-most significant hidden units of the base RBM layer, are expected to consist the first step of disentangling the top-level semantics of the produced generic relations and improve the accuracy of the RBM results.

Further, a modular NN approach will be sought after in order to project the results of the DBNs to natural language, in order to extract the specific semantic associations among related input classes.

## Acknowledgments

This work has been supported by the European Commission under Grant Agreement No. 787061 ANITA.

## References

- [1] M. L. Minsky, Logical versus analogical or symbolic versus connectionist or neat versus scruffy, *AI magazine* 12 (1991) 34–34.
- [2] J. H. Flavell, *Cognitive development.*, Prentice-Hall, 1977.
- [3] W. Huitt, J. Hummel, Piaget’s theory of cognitive development, *Educational psychology interactive* 3 (2003) 1–5.
- [4] G. E. Hinton, A practical guide to training restricted boltzmann machines, in: *Neural networks: Tricks of the trade*, Springer, 2012, pp. 599–619.
- [5] F. Bianchi, P. Hitzler, On the capabilities of logic tensor networks for deductive reasoning., in: *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering*, 2019.
- [6] R. Palm, U. Paquet, O. Winther, Recurrent relational networks, in: *Advances in Neural Information Processing Systems*, 2018, pp. 3368–3378.
- [7] Z. Hu, X. Ma, Z. Liu, E. Hovy, E. Xing, Harnessing deep neural networks with logic rules, *arXiv preprint arXiv:1603.06318* (2016).
- [8] W.-Z. Dai, Q.-L. Xu, Y. Yu, Z.-H. Zhou, Tunneling neural perception and logic reasoning through abductive learning, *arXiv preprint arXiv:1802.01173* (2018).
- [9] M. Noroozi, P. Favaro, Unsupervised learning of visual representations by solving jigsaw puzzles, in: *European Conference on Computer Vision*, Springer, 2016, pp. 69–84.
- [10] G. B. Huang, H. Lee, E. Learned-Miller, Learning hierarchical representations for face verification with convolutional deep belief networks, in: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2012, pp. 2518–2525.
- [11] D. Dligach, T. Miller, Learning patient representations from text, *arXiv preprint arXiv:1805.02096* (2018).
- [12] P. Ristoski, S. Faralli, S. P. Ponzetto, H. Paulheim, Large-scale taxonomy induction using entity and word embeddings, in: *Proceedings of the International Conference on Web Intelligence*, 2017, pp. 81–87.

---

<sup>5</sup>DBNs consist of additional, stacked directed RBM layers over the first undirected RBM layer.

- [13] R. Hu, M. Rohrbach, J. Andreas, T. Darrell, K. Saenko, Modeling relationships in referential expressions with compositional modular networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1115–1124.
- [14] J. Mao, C. Gan, P. Kohli, J. B. Tenenbaum, J. Wu, The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision, arXiv preprint arXiv:1904.12584 (2019).
- [15] L. de Penning, A. Garcez, L. C. Lamb, J. Meyer, A neural-symbolic cognitive agent for online learning and reasoning, in: Proceedings of the Twenty-Second international joint conference on Artificial Intelligence, volume 2, IJCAI, 2011, pp. 1653–1658.
- [16] I. Sutskever, G. E. Hinton, G. W. Taylor, The Recurrent Temporal Restricted Boltzmann Machine, in: Advances in neural information processing systems, 2009, pp. 1601–1608.
- [17] E. De la Rosa, W. Yu, Data-driven fuzzy modeling using Restricted Boltzmann Machines and probability theory, IEEE Transactions on Systems, Man, and Cybernetics: Systems (2018).
- [18] R. Evans, E. Grefenstette, Learning explanatory rules from noisy data, Journal of Artificial Intelligence Research 61 (2018) 1–64.
- [19] G. E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural computation 18 (2006) 1527–1554.
- [20] T. Tieleman, Training Restricted Boltzmann Machines using approximations to the likelihood gradient, in: Proceedings of the 25th international conference on Machine learning, 2008, pp. 1064–1071.
- [21] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, M. Bernstein, L. Fei-Fei, Visual Genome: Connecting language and vision using crowdsourced dense image annotations, 2016. URL: <https://arxiv.org/abs/1602.07332>.
- [22] D. A. Hudson, C. D. Manning, GQA: A new dataset for real-world visual reasoning and compositional question answering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 6700–6709.
- [23] G. A. Miller, WordNet: a lexical database for English, Communications of the ACM 38 (1995).
- [24] S. N. Tran, A. d. Garcez, Knowledge extraction from deep belief networks for images, in: IJCAI-2013 workshop on neural-symbolic learning and reasoning, 2013.
- [25] V. Satopaa, J. Albrecht, D. Irwin, B. Raghavan, Finding a "kneedle" in a haystack: Detecting knee points in system behavior, in: 2011 31st international conference on distributed computing systems workshops, IEEE, 2011, pp. 166–171.
- [26] F. R. Chung, R. P. Langlands, A combinatorial laplacian with vertex weights, journal of combinatorial theory, Series A 75 (1996) 316–327.
- [27] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise., in: Kdd, volume 96, 1996, pp. 226–231.
- [28] P. Kreutzer, G. Dotzler, M. Ring, B. M. Eskofier, M. Philippsen, Automatic clustering of code changes, in: 2016 IEEE/ACM 13th Working Conference on Mining Software Repositories (MSR), IEEE, 2016, pp. 61–72.

**Table 2**

Visual Genome Wordnet synset mismappings.

Wordnet Synset	Sense in VG	Wordnet Synset	Sense in VG
topographic_point.n.01	Spot (as in fleck)	numeral.n.01	Number
contemplation.n.02	Reflection (as in mirroring)	new_jersey.n.01	Jersey (garment)

**Table 3**

GQA RBM and Ground Truth spectral clustering results.

RBM cluster	Common-sense semantics	Ground Truth cluster	Common-sense semantics
people, helmet,	Related to people	cake, spoon, bread, carrot, dish, meat, vegetable, broccoli, tomato, onion, tray, can, sauce, pepper, cheese, label, pizza,	In the kitchen
nose, ear, eye,	On the face	keyboard, computer,	Technological objects
table, plate, bowl, glass,	On a table	orange, apple, fruit, banana, stick,	Fruit
shoe, person, hat, jacket,	What a person wears	outlet, refrigerator, kitchen,	In the kitchen
counter, chair, pillow,	In a room	surfboard, wetsuit,	Related to surfing
shelf, bottle, cabinet,	On the street	bridge, gravel, platform,	Near water
sidewalk, street, sign,	Body parts	bat, jersey, spectator,	Related to baseball game
leg, neck,	Found outdoors (rural)	bird, beak,	Related to bird
fence, field,	Related to animals	rug, carpet, remote control,	In the sitting room
wing, zebra,	Plumbing in room	tag, suitcase,	Related to suitcase
faucet, sink,	What a person wears	brick, balcony,	On/near a balcony
hat, jacket, person,	On a cat		
cat, paw,			

## A. Experimental Results

### A.1. GQA Results

Table 3 portrays the clusters of interrelated classes produced through spectral analysis of an RBM trained on the GQA dataset (objects) and the clusters of most prominently co-occurring classes produced through spectral analysis of the GQA dataset’s ground truth.

### A.2. Visual Genome Results

Table 4 portrays the clusters of interrelated classes produced through spectral analysis of an RBM trained on the Visual Genome dataset (objects) of most prominently co-occurring classes produced through spectral analysis of the Visual Genome dataset’s ground truth respectively.

It is worth mentioning that some Visual Genome object classes are represented with a Wordnet word/term that does not reflect the sense which the Visual Genome data signify. These terms are marked with \* in Table 4 and their proper semantics/senses are listed in Table 2.

**Table 4**

Visual Genome RBM and Ground Truth spectral clustering results.

<b>RBM cluster</b>	<b>Common-sense semantics</b>	<b>Ground Truth cluster</b>	<b>Common-sense semantics</b>
kitchen.n.01, spoon.n.01, bread.n.01, cake.n.03, banana.n.01, pizza.n.01, food.n.01, sauce.n.01, cheese.n.01, plate.n.04, wave.n.01, ocean.n.01, surfboard.n.01, beach.n.01, telephone.n.01, root.n.03,	Related to food	paw.n.01, bear.n.01, fur.n.01,	Related to bear
vegetable.n.01,	Things at the beach	ring.n.01, finger.n.01,	On a hand
room.n.01, drawer.n.01, cabinet.n.01,	Related to vegetation	sauce.n.01, bread.n.01, tomato.n.01, vegetable.n.01, meat.n.01, cheese.n.01, fork.n.01, napkin.n.01, spoon.n.01, tray.n.01, pizza.n.01, knife.n.01, food.n.01, banana.n.01, cake.n.03, root.n.03,	Related to food
bird.n.01, beak.n.01,	Things in a room	court.n.01, racket.n.04, player.n.01, skirt.n.01,	Related to tennis
sink.n.01, faucet.n.01, bathroom.n.01,	Related to birds	bat.n.05, new_jersey.n.01*, uniform.n.01,	Related to sports/ baseball
uniform.n.01, ball.n.01, new_jersey.n.01*, player.n.01, bat.n.05, word.n.01, writing.n.01, sign.n.02,	Things in a bathroom	sink.n.01, faucet.n.01, bathroom.n.01, toilet.n.01,	Things in a bathroom
arrow.n.01, traffic_light.n.01,	Related to sports/baseball	animal.n.01, cow.n.01, sheep.n.01,	Related to farm animals
computer.n.01, laptop.n.01, screen.n.01, television.n.01, wire.n.01, keyboard.n.01, pillow.n.01, desk.n.01, bed.n.01, goggles.n.01, board.n.02, ski.n.01,	Related to writing	kitchen.n.01, drawer.n.01,	Things in a kitchen
sand.n.01, mountain.n.01, hill.n.01, path.n.04,	Related to traffic signs	beak.n.01, bird.n.01,	Related to birds
soil.n.02, shrub.n.01, flower.n.01, plant.n.01, cow.n.01, sheep.n.01,	Things in a sitting room	hoof.n.01, zebra.n.01, mane.n.01, horn.n.01,	Related to animals
napkin.n.01, rug.n.01, tray.n.01, shelf.n.01, ceiling.n.01, dress.n.01, skirt.n.01, lady.n.01, child.n.01, bag.n.04, male_child.n.01, face.n.01, topographic_point.n.01*, hoof.n.01, tail.n.01,	Related to snow sports	computer.n.01, keyboard.n.01, laptop.n.01, desk.n.01, screen.n.01, television.n.01,	Things in a sitting/study room
branch.n.01, leaf.n.01,	Related to path	skateboard.n.01, ramp.n.01,	Related to skateboard
people.n.01, numeral.n.01*,	Related to vegetation	traffic_light.n.01, arrow.n.01, license_plate.n.01,	On the street
windshield.n.01, headlight.n.01,	Farm animals	ski.n.01, goggles.n.01,	Related to snow sports
license_plate.n.01, motorcycle.n.01,	Inside eating area		
wing.n.01, airplane.n.01,	Related to lady		
fork.n.01, tomato.n.01, sidewalk.n.01, field.n.01, umbrella.n.01, contemplation.n.02*, bridge.n.01, logo.n.01, letter.n.01, design.n.01,	Facial attributes		
ring.n.01, finger.n.01, paper.n.01, box.n.01,	Animal attributes		
	Related to tree/shrub		
	Related to population		
	On a car/motorcycle		
	Related to motorcycle		
	Related to airplane		
	Related to food		
	Surfaces		
	Related to water		
	Related to logo/posters		
	On a hand		
	Related to paper		