Learning Plausible Inferences from Semantic Web Knowledge by Combining Analogical Generalization with Structured Logistic Regression

Chen Liang & Kenneth D. Forbus
Qualitative Reasoning Group, Northwestern University
2133 Sheridan Road, Evanston, IL, 60208 USA
contact: chenliang2013@u.northwestern.edu

Abstract

Fast and efficient learning over large bodies of commonsense knowledge is a key requirement for cognitive systems. Semantic web knowledge bases provide an important new resource of ground facts from which plausible inferences can be learned. This paper applies structured logistic regression with analogical generalization (SLogAn) to make use of structural as well as statistical information to achieve rapid and robust learning. SLogAn achieves state-of-the-art performance in a standard triplet classification task on two data sets and, in addition, can provide understandable explanations for its answers.

Introduction

Learning is a key part of cognitive systems. Humans are fast, efficient learners: with a few observations, we can acquire new knowledge to reason and make predictions. Moreover, we can also explain our reasoning to others to convince or persuade them. These two abilities are the focus of this work. Analogical learning has proven to be a promising model for human learning. It has been used to learn spatial prepositions (Lockwood et al. 2008), learn sketched concepts (McLure et al 2015), and construct hierarchical concepts (Liang & Forbus, 2014). Computational models of structural alignment and structure mapping have demonstrated the ability to learn linguistic patterns at a similar pace to human infants. (Kuehne et al. 2000). SAGE, which implements analogical generalization, is used as a learning module for the Companion cognitive architecture (Forbus et al. 2009).

Like humans, cognitive systems need knowledge and the ability to reason with it. Hand-coding knowledge and inference rules is not a scalable solution. Semantic web knowledge bases are a potential resource for cognitive systems to automatically acquire common sense knowledge and learn plausible inferences. Semantic web KBs like Freebase, WordNet, and YAGO have accumulated considerable structured data, which is already used to support AI tasks like question answering and information retrieval. These KBs continue to grow rapidly.

Using semantic web KBs brings up two concerns. First, because they can be extracted from text or collected by crowd-sourcing, they are often incomplete and noisy. Traditional logical inference may not be sufficiently robust to reason over them. Second, unlike images, auditory data or raw text, semantic web data is inherently structured. Although statistical methods can handle uncertainty well, most such methods are designed to work over feature vectors, and are not able to operate over structures.

Statistical relational learning focuses on extending statistical machine learning methods from feature vectors to relational data. For example, Markov logic networks (Richardson & Domingos, 2006) combine first-order logic with Markov networks, providing the representation power of the former and the statistical power of the latter. But it suffers from scalability problems. Other models, like neural networks or bilinear models, work on vector embedding of entities (Socher et al. 2013, Wang et al. 2014). They can achieve high prediction accuracy efficiently, but the opacity of the models makes it hard for users to interpret and understand their answers.

This paper describes how structured logistic regression with analogical generalization (SLogAn) can be used to learn plausible inferences from semantic web knowledge. We start by summarizing the models we build upon, and then describe our new method, covering case construction, learning, prediction, and explanation generation. We compare it to state-of-the-art methods on a standard triplet classification task with two datasets, showing that it has competitive performance, and, at the same time, can provide understandable explanations for its answers.

Background

We assume Gentner's (1983) structure-mapping theory. Our model is built upon the Sequential Analogical Generalization Engine (SAGE; McLure et al. 2010), which in turn uses the Structure-Mapping Engine (Falkenhainer et al. 1989) for analogical comparison and MAC/FAC (Forbus et al. 1995) for analogical retrieval. We start with SME since it is the most fundamental. SME takes as input two structured representations, a *base* and *target*, and produces one or more mappings. Each mapping provides a set of correspondences (i.e. what goes with what), a structural evaluation score which provides an overall estimate of match quality, and candidate inferences. We refer to the similarity score of a mapping as NSIM(base,target), which is normalized to [0,1] by dividing the raw score by the mean of the self-scores of the base and target[1]. Forward candidate inferences go from base to target, reverse candidate inferences go from target to base. MAC/FAC takes as input a case library, which is a set of structured descriptions, and a probe, which is a structured description. It returns one or more approximations to the most similar case in the case library, using a two-stage process that enables it to scale to large case libraries. The first stage uses a flattened version of the relational structure of cases, called *content vectors*, whose dimensions are proportional to the weighted number of occurrences of each predicate in a description. The dot product of two content vectors is an estimate of SME's structural evaluation score for the structured representations, making it a useful coarse filter. Both SME and MAC/FAC have been used to model a variety of psychological phenomena.

SAGE maintains, for each concept, a *generalization context*. A generalization context has a trigger, which is used to test whether or not an incoming example should be added to it. (An incoming example might satisfy multiple triggers, and hence be processed by several generalization contexts.) Each generalization context maintains a set of *generalizations* and a set of *unassimilated examples*. (Either of these sets might be empty, and both are initially.) Generalizations are also structured representations, but associated with their statements are probabilities, based on the number of times facts that align with them are found in examples that are part of that generalization.

Every time a new example is added, SAGE uses MAC/FAC to retrieve up to three examples or generalizations, based on whatever is the most similar to the new example. If nothing is retrieved, or the similarity to the returned item is less than an *assimilation threshold*, the new example is stored as is. Otherwise, if the returned item is a generalization, the new example is assimilated into it. If the returned item is a previously unassimilated example, then the two are combined into a new generalization.

The assimilation process increments frequency counts associated with each statement, based on whether or not something in the example aligned with it. For a new generalization, such facts are always either 1.0 (in both) or 0.5. If, for example, one black cat and two grey cats had been seen, then P[(primaryObjectColor <*GenEnt*> Black)] = 1/3. Facts whose probabilities drop too low are pruned, for efficiency. Importantly, these generalizations do not have logical variables: When non-identical entities are aligned, as in the cats example, a new arbitrary individual (called <*GenEnt*> above) is constructed to stand for the aligned individuals, with its characteristics being determined by the set of statements in the generalization that constrain it.

In most semantic web KBs, knowledge is stored in the triplet format: "entity relation entity". Since there are only binary relations, it can be seen as a labeled, directed graph. Each entity is a node, and each triplet between two entities is an edge labeled with the relation. Semantic web KBs have been used in a variety of tasks, including triplet classification. Given triplets extracted from a KB, the system should learn to distinguish correct triplets like "Obama nationality USA" from incorrect ones like "Obama nationality Kenya". Then, the learned model is tested on a holdout test set, with performance measured by classification accuracy.

Method

We extend analogical generalization with structured logistic regression to make use of the structural as well as the statistical information in the semantic web KB. The goal is to learn what kinds of inferences are plausible. For example, when learning about relation "nationality", we are not learning what people from different countries look like. Instead, we are learning about how to infer a person's nationality. One possible inference might be "this person comes from Chicago, and I know several people from Chicago. What are their nationalities?"

Case construction by path-finding

The first step is to prepare the input data. Analogical comparison and generalization work on cases. For this task, cases contain structured information about particular triplets. The idea is to include enough information to enable the

[1] The mapped representations are subsets of both base and target, so its score is lower than either of their self-scores.

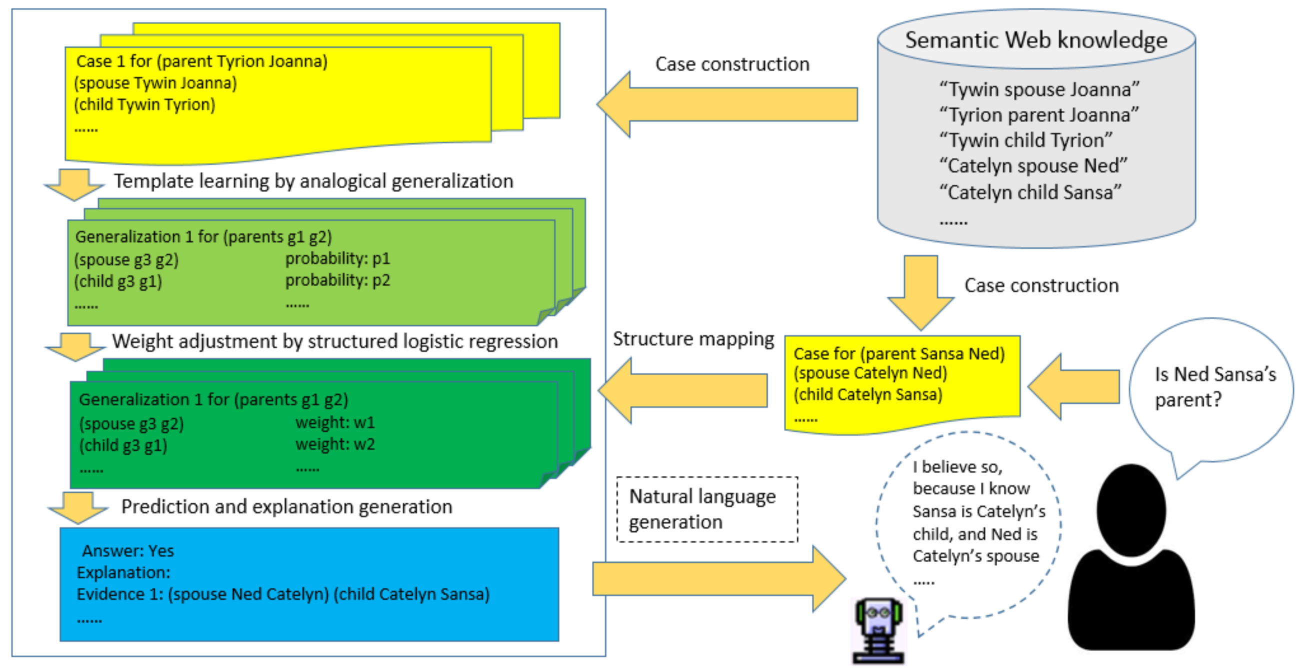

Figure 1: Intended workflow for SLogAn. The learning system gets input from semantic web knowledge base, does template learning and weight adjustment with the data. Using the learned generalizations, it provides answers, with explanations, for the user's queries. Natural language generation (dashed lines) is not yet implemented.

system to distinguish correct from incorrect triplets, while limiting the size of the cases for the sake of efficiency and to reduce the number of irrelevant matches.

As discussed above, a semantic web KB can be viewed as a graph. We use path-finding as a heuristic to pick relevant facts for case construction. For example, when we want to create a case for a triplet with A and B as head and tail entities, we use depth-first-search to find paths between them and put all the facts along the paths found into the case. In a large scale and highly connected knowledge base, an exhaustive search will be prohibitive, so we use limits on branching factor and search depth to randomly select parts of the search tree to explore.

These two thresholds control the tradeoff between information and efficiency. The larger the branching factor, the more information is in the case, and the larger the search depth, the more distant are the relationships between the two entities that can be found. The bound on branching factor is applied to each type of relation separately. For example, if we hit the entity "Chicago" in the search, and it has a "location" relation with 100 people and a "place_of_birth" relation with 10 people and the upper limit is 20, we randomly select 20 of the 100 Chicagoans to explore, and use all of the 10 people born in Chicago to explore, because the number is less than 20.

The intuition is that the facts we use to infer a relation will be represented by a small number of fixed paths between the entities in the case. In other words, if some relation holds between two entities, they are likely to be indirectly related to each other in some other ways.

Since there are only positive examples in the original dataset, we corrupt the correct triplets by changing their tail entities to wrong ones to get negative examples, just like Socher et al. (2013) did. We treat the training set as our knowledge base, and create cases from it. With just a few cases, SLogAn is able to learn plausible inferences. For the triplet classification task, we used only 10 positive and 10 negative examples for each relation.

Analogical generalization: structural alignment for template learning

The second step is to decide what to include in the inference, i.e., learning a template for the inference. This is done by SAGE. As outlined above, SAGE can create generalizations by comparing examples and compressing them into one prototype. A generalization works as a template for an inference. It is trivial to compare feature vectors because they all share the same dimensions, but it takes some effort to find the best way to align structured representations of examples and compress them into one structurally consistent template that summarizes the facts. SME does this job by finding the best structural alignment with systematicity as main criterion and several constraints to ensure structural consistency. We also require the head entity and tail entity of one triplet to be respectively matched to those of the other triplet in the mapping to make sure they are the focus of the template. Note that, unlike the way

Explanation for (taufaahau_tupou_iv ethnicity tongans):
Evidence 1: (taufaahau_tupou_iv parents viliami_tungi_mailefihi) (viliami_tungi_mailefihi ethnicity tongans)
Evidence 2: (viliami_tungi_mailefihi children taufaahau_tupou_iv) (viliami_tungi_mailefihi ethnicity tongans)
Evidence 3: (taufaahau_tupou_iv nationality tonga) (george_tupou_i_of_tonga nationality tonga) (george_tupou_i_of_tonga ethnicity tongans)
comment: I believe Taufa'ahau Tupou IV's ethnicity is Tongan, because I know his parent's ethnicity is Tongan, and I remember a person from the same country as him, whose ethnicity is also Tongan.

Explanation for (qusay_hussein parents saddam_hussein):
Evidence 1: (qusay_hussein parents sajida_talfah) (uday_hussein parents sajida_talfah) (uday_hussein parents saddam_hussein)
Evidence 2: (qusay_hussein parents sajida_talfah) (uday_hussein parents sajida_talfah) (saddam_hussein children uday_hussein)
Evidence 3: (qusay_hussein parents sajida_talfah) (saddam_hussein spouse sajida_talfah)
comment: I believe Saddam is Qusay's parent, because I know Saddam is Qusay's sibling Uday's parent and Qusay's parent Sajida's spouse.

Explanation for (vuk_stefanovic_karadzic religion serbian_orthodox_church):
Evidence 1: (vuk_stefanovic_karadzic ethnicity serbs) (gavrilo_princip ethnicity serbs) (gavrilo_princip religion serbian_orthodox_church)
Evidence 2: (vuk_stefanovic_karadzic ethnicity serbs) (zoran_in_ic ethnicity serbs) (zoran_in_ic religion serbian_orthodox_church)
Evidence 3: (vuk_stefanovic_karadzic ethnicity serbs) (alexander_i_of_yugoslavia ethnicity serbs) (alexander_i_of_yugoslavia religion serbian_orthodox_church)
comment: I believe Vuk Stefanovic Karadzic's religion is the Serbian orthodox church, because I can recall several persons with the same ethnicity, whose religion is the Serbian orthodox church.

Figure 2: Examples of explanations and comments: with only ground facts and no prior knowledge about the relations, the system learns to make inference about them with high accuracy and provide the explanations for its answers. The comments are generated manually.

SAGE is usually used, we first created generalizations with positive examples only, and then add negative examples to these generalizations so that they contain facts from positive as well as negative examples. In this way, some facts could contribute negatively to the target relation. For example, if I know that B is A's parent, then A cannot be B's parent. Although that fact never appears in a positive example, it is still a critical fact to consider in the inference.

Using multiple generalizations together as one template has strong expressive power. Learning plausible inferences is similar to learning the concept of the target relation. A concept can be represented by several generalizations in a generalization context, which is analogous to a disjunctive normal form. Each generalization can be seen as a conjunction of facts inside it. Facts with negative weights play the role of negations. Combining multiple generalizations forms a disjunctive representation of the concept. If we also combine the generalizations with structured logistic regression, the model can be seen as a structured 2-layer neural network. Each generalization plays the role of a logistic unit, but the weight vector of the unit in standard neural network is changed to a structured template with associated weights to handle structured input instead of feature vectors. Correspondingly, the dot-product operation to activate a unit in standard neural network is replaced with a structure mapping process. In this triplet classification task, although multiple generalizations can help in accuracy, we did not use them for the sake of efficiency.

Structured logistic regression: structure mapping and gradient descent for weight adjustment

The third step is to learn how much each fact in the template supports or contradicts the target relation. Each expression in the template is associated with a weight measuring its positive support. So if a fact contradicts the target relation, it gets a negative score. We use the probability of each fact as an initial value for its weight. To further adjust the weights, we propose *structured logistic regression*, which is an extension of logistic regression from feature vectors to structured cases by combining it with structure mapping.

Structured logistic regression works as follows. Given an example, we compare it to the generalization with SME. With the resulting mapping, we compute the prediction score of the example being a positive example with

$$p = \frac{1}{1 + e^{-s}}$$

where s is the similarity score computed by SME. Then, we use cross entropy to define the prediction error and L1 regularization to promote sparsity in weights for better explanation generation:

$$J = -\sum_{i=0}^{n} [(1 - y_i) \log(1 - p_i) + y_i \log(p_i)] + \alpha \sum_{j=1}^{m} |w_j|$$

n and m are the number of examples and facts in the generalizations respectively. y_i and p_i are the label (1 for positive, 0 for negative) and prediction score of the ith example. α controls the strength of the regularization. Since this error depends on similarity score and similarity score depends on the weights, we can calculate the derivate of the error with respect to the weights and do gradient descent on them. The prediction error on the validation set is used to decide when to stop the weight adjustment.

Prediction and explanation generation by finding most weighted paths

After training, the model has learned to assign high scores to the positive examples and low scores to the negative examples. We use the validation set to decide the criterion c. When the score is larger than c, the model predicts true, otherwise it predicts false.

Explanations are representations of one's own reasoning which can be understood by others. Since SLogAn uses the structured representation directly, it is more interpretable for the user. By finding the highest weighted paths, we could provide understandable explanations. The intuition here is that each path between two entities represents one inference chain. From the weights, we can know whether a given inference chain supports or contradicts the target relation, and how important it is. Given a path between the two entities in the query triplet, we use the average weight of the facts along this path as its weight. Then, we rank the paths with the absolute value of their weights. Finally, we pick several paths on the top as the explanation for the answer. To make the explanation clearer to the user, we use L1 regularization to induce sparsity in the weight adjustment, so that it prefers a few high weights rather than many low weights, in other words, it provides a few strong pieces of evidence rather than many weak pieces of evidence.

Experiment

To test it, we compare SLogAn's performance to state-of-the-art performance on the triplet classification task. The datasets were collected by Socher et al. (2013). One dataset contains triplets from Freebase consisting of 13 relations and 38,696 entities, the other dataset, from WordNet, consisting of 11 relations and 75,043 entities. More information about the number of triplets in training, validation and test sets is shown in Table 1.

Datasets	#Relations	#Entities	#Train	#Validation	#Test
WN11	11	38,696	112,581	2,609	10,544
FB13	13	75,043	316,232	5,908	23,733

Table 1: Statistics of WordNet and Freebase datasets

As for the hyperparameters, the limit on branching factor is 20 and the search depth is 3 for FB13 and 5 for WN11, which is chosen based on run time limit. (On average, one entity in FB13 has $316232/75043 \approx 4.2$ outgoing triplets, while the same number is $112581/38696 \approx 2.9$ in WN11, so the search depth is lower on FB13.) For each relation, we randomly select 10 triplets as positive examples, and create their corresponding negative examples as training data. The regularization parameter α is set to 0.01, which is decided by prediction error on validation set.

Other methods have been tested on exactly the same datasets, making the results useful for comparison. Table 2 shows accuracy performances from Wang et al (2014). SLogAn is 2nd on WN11 and FB13, and no method consistently outperforms it on both datasets. Thus our method performs at a level that is competitive with the state-of-the-art. Other approaches have not provided information concerning statistical significance, so we will not be able to compare that against them. The difference between SLogAn and chance is statistically significant with $p < 0.001$. Prediction accuracies of different relations are shown in Table 3. SLogAn learns from the data collected by path-finding during case construction, so its performance drops on relations like "similar to" because it is hard to find a path in the KB between entities like "adamant_1" and "physiologist_1", although they are "similar to" each other.

Model	WordNet11	Freebase13
Distance Model	53.0	75.2
Hadamard Model	70.0	63.7
Single Layer Model	69.9	85.3
Bilinear Model	73.8	84.3
Neural Tensor Network	70.4	87.1
TransH	78.8	83.3
SLogAn	75.3	85.3

Table 2: Triplet classification accuracy (%) of different models

The results provided by other methods are opaque to users. Since they are based on vector embedding, they cannot easily explain why they make certain predictions. In contrast, we work on the structured data directly. From the explanation the model provides, the user is able to see how its reasoning works. Examples of explanations are shown in Figure 2. Note that unfamiliar examples are intentionally chosen to show how explanations help with human

WordNet11	accuracy	Freebase13	accuracy
has instance	74.5	gender	85.9
type of	76.9	nationality	94.6
member meronym	75.8	profession	81.4
member holonym	74.2	institution	79.4
part of	72.0	cause of death	76.9
has part	73.5	religion	82.2
subordinate instance of	81.7	ethnicity	87.9
domain region	69.9		
synset domain topic	76.0		
similar to	50.0		
domain topic	69.0		

Table 3: Triplet classification accuracy (%) of different relations

validation of the results and increase trust in prediction: even without explanations, familiar examples are easy for people to validate. We used the top 3 most weighted paths as the explanation for each example.

Discussion

SLogAn learned plausible inferences from semantic web knowledge to achieve competitive performance to state-of-the-art methods on the triplet classification task, and, moreover, provides understandable explanations. These two abilities would make a cognitive system more autonomous and independent because it does not need to be told about everything it will be asked for, and more trustworthy and helpful because it can explain its answer to the user.

For example, when asked about "is Taufa'ahau Tupou IV's ethnicity Tongan?", even though the system is not told about this fact or how to infer one's ethnicity, it can learn it from several examples it knows of the relation "ethnicity" and makes a prediction based on its knowledge of Taufa'ahau Tupou IV and Tongan with reasonable accuracy. If the users are only provided with the answer, they have to decide whether to believe it or not based on their trust of the system. However, if the system could say, "I believe Taufa'ahau Tupou IV's ethnicity is Tongan, because I know his parent's ethnicity is Tongan, and I remember a person with the same nationality whose ethnicity is also Tongan", it is easier to convince the users when the system is correct, and alert them when the system is wrong. It would make cooperation more efficient and natural, by reducing the labor of human validation.

Related work

Many vector embedding based methods have been tested on the triplet classification tasks. Socher et al. (2013) and Wang et al. (2014) are the most recent and have the best performance. Their models are quite different from ours. They learn vector representations for the entities and parameterization (translation or tensor) for each relation that implicitly encode the knowledge about them. During training, they created negative examples for every triplet in the training set and train on all of them. In contrast, SLogAn only randomly selects a few triplets and creates cases and corresponding negative examples for them. Instead of implicit encoding of knowledge with vector embedding, we use the structured knowledge directly and learn generalizations that explicitly encode valid inferences. This gives our model interpretability. Moreover, SLogAn can learn each relation separately, while theirs have to learn all the relations together.

Several previous efforts have used path-finding for relational learning. Sharma & Forbus (2010) used higher-order knowledge about relations and reinforcement learning to construct plausible inference patterns, whereas SLogAn is learning from purely ground facts. Integrating these methods could prove valuable. Richards & Mooney (1992) uses path-finding to find candidate clauses for learning first-order rules, but it was computationally expensive. Lao et al. (2011) used limited-length path-finding in NELL knowledge base to create features and do logistic regression with them. As is discussed above, SLogAn can use multiple generalizations to handle disjunctive concepts, and thus has more expressive power. Also note that SLogAn can potentially work with other case construction methods as long as they provide structured representations of relevant information.

Structural logistic regression (Popescul & Ungar, 2003) generates features by propositionalizing first-order rules learned by inductive logic programming, and uses logistic regression with these features for classification. Relational logistic regression (Kazemi et al. 2014) uses logistic regression to learn weights for first-order formulae in defining the conditional probability of a new relation given those formulae. Their ways of adding the counts of certain facts as features are possible improvements for the current model. Although the triplet classification task only deals with binary relations, SLogAn has the ability to deal with higher arity and high-order relations because it builds on SME and SAGE, which can handle them. Halstead & Forbus (2005) takes a similar approach to this work. They used SAGE to generate probabilistic generalizations to turn structured cases into features and built a Bayesian network on top of them to make predictions.

Deductive systems like Cyc and SHAKEN (Clark et al. 2001) can also provide explanations, but those explanations tend to be narrow and deeply nested, whereas the explanations produced by our current system are broad and shallow, more akin to explanations found in evidential reasoning arenas (e.g. intelligence analysis). Extensions to handle deeper nested inference by adding more levels to the current system are interesting, and will be crucial for learning more complex yet still interpretable models for prediction.

Conclusion and future work

This work shows that analogical generalization and structure mapping can be combined with statistical machine learning methods to achieve state-of-the-art performance on a standard task, while preserving interpretability. It enables cognitive systems to learn from structured data, e.g. semantic web resources, to do reasoning with what they have learned, and to explain their reasoning to the user.

We plan to explore several future directions. First, semantic web knowledge typically contains only binary relations involving entities. However, SME and SAGE are designed to work on cases that also include high-order relations that express arguments and explanations. We plan to test this method with representations produced via learning by reading systems (Forbus et al. 2007) and sketch understanding systems (Forbus et al. 2011). Second, we plan to compare its performance to other SRL methods on other tasks, such as link prediction. Third, unsupervised learning of high-level representations to disentangle independent factors and discover patterns has recently drawn a lot of attention (Lee et al. 2011, Bengio, 2009). The learned representations can be used to represent the input more compactly and help supervised learning to achieve better performance. Extending the current model to learn high-level representations for structured examples that are understandable, more compact and useful for supervised learning is a promising direction. Fourth, we plan to integrate it into our Companion cognitive architecture, to provide a new learning capability and exploit Companion's interaction capabilities to implement the full workflow of Figure 1.

Acknowledgements

This research was sponsored by the Socio-Cognitive Architectures Program of the Office of Naval Research, N00014-13-1-0470.

References

Bengio, Y. (2009). Learning Deep Architectures for AI. *Found. Trends Mach. Learn.*, 2(1), 1-127. doi: 10.1561/2200000006

Clark, P. et al. (2001). Knowledge entry as the graphical assembly of components. *Proceedings of the 1st international conference on Knowledge capture* (pp. 22-29). ACM.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial intelligence*, 41(1), 1-63.

Forbus, K., Riesbeck, C., Birnbaum, L., Livingston, K., Sharma, A., & Ureel, L. (2007). Integrating Natural Language, Knowledge Representation and Reasoning, and Analogical Processing to Learn by Reading. *Proceedings of AAAI-07*.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). CogSketch: Sketch Understanding for Cognitive Science Research and for Education. *Topics in Cognitive Science*, 3(4), 648-666. doi: 10.1111/j.1756-8765.2011.01149.x

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A Model of Similarity-Based Retrieval. *Cognitive Science*, 19(2), 141-205.

Forbus, K. D., Klenk, M., & Hinrichs, T. (2009). Companion cognitive systems: Design goals and lessons learned so far. *IEEE Intelligent systems*, 24(4), 36-46.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, 7(2), 155-170.

Halstead, D. and Forbus, K. (2005). Transforming between Propositions and Features: Bridging the Gap. *Proceedings of AAAI-2005*. Pittsburgh, PA.

Kazemi, S. M., Buchman, D., Kersting, K., Natarajan, S., & Poole, D. (2014). Relational logistic regression. *Proc. 14th International Conference on Principles of Knowledge Representation and Reasoning (KR-2014)*.

Kuehne, S., Gentner, D. and Forbus, K. (2000). Modeling infant learning via symbolic structural alignment. *Proceedings of CogSci 2000, August.*

Lao, N., Mitchell, T., & Cohen, W. W. (2011). Random walk inference and learning in a large scale knowledge base. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, United Kingdom.

Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2011). Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Communications of the ACM*, 54(10), 95-103.

Liang, C. and Forbus, K. (2014). Constructing Hierarchical Concepts via Analogical Generalization. *Proceedings of the Cognitive Science Society.*

Lockwood, K., Lovett, A., & Forbus, K. (2008). Automatic classification of containment and support spatial relations in English and Dutch. In *Spatial Cognition VI. Learning, Reasoning, and Talking about Space* (pp. 283-294). Springer Berlin Heidelberg.

McLure, M., Friedman, S., & Forbus, K. (2015). Extending Analogical Generalization with Near-Misses. *Proceedings of AAAI-2015.*

McLure, M., Friedman, S., & Forbus, K. (2010). Learning concepts from sketches via analogical generalization and near-misses. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society (CogSci)*, Portland, OR.

Popescul, A., & Ungar, L. H. (2003). Structural logistic regression for link analysis. Departmental Papers (CIS), 133.

Richards, B. L., & Mooney, R. J. (1992). Learning relations by pathfinding. *Proceedings of the tenth national conference on Artificial intelligence.*

Richardson, M., & Domingos, P. (2006). Markov logic networks. *Machine learning*, 62(1-2), 107-136.

Sharma, A. and Forbus, K.D. (2013). Graph Traversal Methods for Reasoning in Large Knowledge-Based Systems. *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI-13)*, Bellevue, Washington.

Socher, R., Chen, D., Manning, C. D., & Ng, A. (2013). Reasoning With Neural Tensor Networks for Knowledge Base Completion. *Proceedings of Advances in neural information processing systems*.

Wang, Z., Zhang, J., Feng, J., & Chen, Z. (2014). Knowledge Graph Embedding by Translating on Hyperplanes. *Proceedings of Twenty-Eighth AAAI Conference on Artificial Intelligence.*