Reasoning about Unmodelled Concepts — Incorporating Class Taxonomies in Probabilistic Relational Models

Daniel Nyga and Michael Beetz
Institute for Artificial Intelligence
University of Bremen, Germany
{nyga, beetz}@cs.uni-bremen.de

Abstract

A key problem in the application of first-order probabilistic methods is the enormous size of graphical models they imply. The size results from the possible worlds that can be generated by a domain of objects and relations. One of the reasons for this explosion is that so far the approaches do not sufficiently exploit the structure and similarity of possible worlds in order to encode the models more compactly. We propose fuzzy inference in Markov logic networks, which enables the use of taxonomic knowledge as a source of imposing structure onto possible worlds. We show that by exploiting this structure, probability distributions can be represented more compactly and that the reasoning systems become capable of reasoning about concepts not contained in the probabilistic knowledge base.

Introduction

Many real-world reasoning problems require the combination of relational representations with inference mechanisms that can solve the problems by reasoning from incomplete, ambiguous, inaccurate and even contradictory information. Examples of such reasoning tasks are the interpretation of natural-language [Beltagy & Mooney, 2014], object recognition for robot perception [Nyga, Balint-Benczedi, & Beetz, 2014] or intent recognition in human-robot interaction [Sukthankar *et al.*, 2014].

First-order probabilistic models [Getoor *et al.*, 2007] have great potential to serve as powerful problem-solving tools for such application domains: joint probability distributions over the instantiated relations that describe the possible worlds in the respective domain can be queried for any aspect Q contained in the model given any evidence E, $P(Q \mid E)$.

These powerful reasoning capabilities, however, come at the cost of computational complexity in learning and reasoning as the size of the domain under consideration grows. As a consequence, practical applications are mostly bound to small application domains with limited complexity. Many knowledge systems, however, have to work in open worlds: they are equipped with knowledge bases (KB) that have to answer queries about unseen situations that have not been accounted in their design, such as the examples mentioned above.

Hence, the application of expressive probabilistic representation methods requires the inference mechanisms to support *off-domain reasoning* – reasoning about concepts that are not explicitly represented in the KB. Most of the probabilistic models, however, do not support off-domain reasoning. They require every symbol subject to reasoning to be explicitly represented. On the other hand, learning a probability distribution with all possible concepts is hopelessly infeasible.

We therefore aim at developing reasoning mechanisms that are able to rapidly yet flexibly generalize and learn from very few examples, which has also been identified as key features in human cognition [Tenenbaum *et al.*, 2011; Bailey, 1997]. An obvious idea to tackle this is to take into account knowledge about the taxonomic structure of the reasoning domain, which is captured by ontological knowledge representations such as description logics. To this end, the correlation between the semantic similarity of concepts, and the similarity of their relational structure can be exploited for reasoning in probabilistic relational models to transfer the learned knowledge to classes unseen in the training data.

We propose FUZZY-MLNs as a probabilistic reasoning framework for Markov logic networks (MLN) [Richardson & Domingos, 2006]. FUZZY-MLNs exploit the semantic similarity of concepts in a taxonomy in order to handle off-domain concepts in previously unseen situations in a meaningful way and hence allow efficient generalization from very sparse data whilst the original representation formalism of MLNs remains unchanged. The key idea of FUZZY-MLNs is to learn joint probability distributions *conditioned on* large taxonomic knowledge bases that are assumed to be given as factual knowledge. Indeed, a number of comprehensive high-quality taxonomies exist that have been carefully designed to reflect the semantic similarity of concepts [Fellbaum, 1998; Lenat, 1995], which we use as an implementational basis. In contrast to existing probabilistic methods incorporating class hierarchies, FUZZY-MLNs do not target reasoning about the taxonomic structure as such. This comes with the advantage that the concepts subject to reasoning do not need to be exhaustively modelled in the probabilistic KB. This enables (1) compact representation of knowledge, (2) powerful generalization from sparse training data and (3) reduced complexity of learning and inference. In particular, our contributions are the following:

1. We present an approach for reasoning about unknown concepts by exploiting semantic similarity to known concepts in Markov logic, which typically impedes a compact representation of classes that are hierarchically organized in a taxonomy.
2. We propose a reasoning framework for MLNs that enables inference in presence of vague evidence, which allows a very compact representation of knowledge in MLNs and learning from sparse data.
3. We demonstrate the strengths of FUZZY-MLNs by the example of word-sense disambiguation and showcase its strong generalization abilities.

Running Example

Let word-sense disambiguation (WSD) and semantic role labelling (SRL), which are widely studied problems in natural-language processing, be our running examples. Solving these problems enables software systems to interpret incomplete and ambiguous instructions and transform them into well-defined action specifications. More specifically, take the terms ‘cup’ and ‘milk’ and their usage in the two instructions ‘fill a cup with milk’ and ‘add a cup of milk’. In the former case, ‘cup’ refers to a drinking mug, a physical object that can hold milk. In the latter case, it refers to a measurement unit specifying the amount of milk to be added to something not further specified.

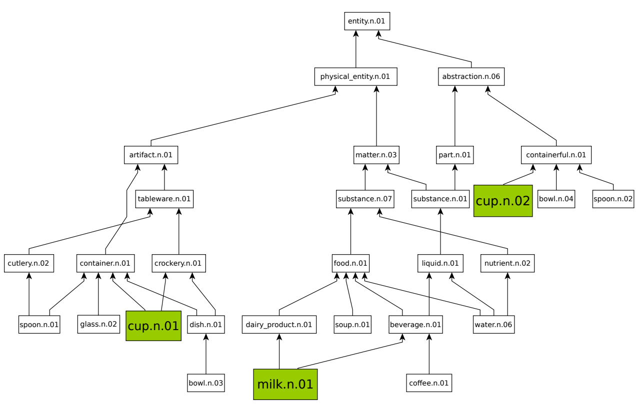


Figure 1: Excerpt of the WordNet taxonomy of concepts for the ‘containers-&-liquids’ example.

Figure 1 shows a small excerpt of the WordNet taxonomy of possible word senses covering this example. Using the taxonomy we can represent the two instructions using the following logical assertions

instruction 1:	**instruction 2:**
instance-of(*Fill*, *fill-sense*)	*instance-of*(*Add*, *add-sense*)
is-a(*fill-sense*, *fill.v.01*)	*is-a*(*add-sense*, *add.v.01*)
instance-of(*cup*, $cup\text{-}sense_1$)	*instance-of*(*cup*, $cup\text{-}sense_2$)
is-a($cup\text{-}sense_1$, *cup.n.01*)	*is-a*($cup\text{-}sense_2$, *cup.n.02*)
instance-of(*milk*, *milk-sense*)	*instance-of*(*milk*, *milk-sense*)
is-a(*milk-sense*, *milk.n.01*)	*is-a*(*milk-sense*, *milk.n.01*)
sem_role(*cup*, *goal*)	*sem_role*(*cup*, *amount*)
sem_role(*milk*, *theme*)	*sem_role*(*milk*, *theme*),

The assertions assign a word sense (*instance-of*) to each word. The word sense is linked to the taxonomy using the *is-a* predicate. In addition, the predicate *sem_role* states the semantic role that the word takes in the instruction, whether it is the object acted on, the source of the stuff to be transferred, the destination, the action verb, and so on.

Now suppose we have a taxonomy and two examplary instructions to learn from: ‘fill a cup with milk’ and ‘add a cup of milk’. For the sentence ’fill water into the pot’ a probabilistic reasoner should infer that water is the stuff to be added and the pot the destination, even when ’water’ and ’pot’ are not contained in the probabilistic knowledge base. The reason is that ’water’ is a liquid like ’milk’ and therefore semantically similar and that a ’pot’ is also a container and therefore similar to a cup. Current first-order probabilistic reasoning frameworks cannot perform this pattern of reasoning as they are restricted to concepts contained in their probabilistic knowledge base.

In the following sections we will explain how we can extend MLNs to perform such reasoning tasks. Note that the reasoning tasks we are interested in are not whether or not two concepts are similar. This is already asserted in the taxonomy. We rather want to infer the concepts that entities belong to and the role they take in actions.

Foundations

Before defining FUZZY-MLNs we first introduce the formal groundwork they are based on: *Description Logics* (DL), *Fuzzy Logics* (FL) and *Markov logic networks* (MLN).

Markov Logic Networks Our basic formalism for representing, learning, and reasoning about first-order probabilistic knowledge bases are MLNs. Formally, an MLN L is given by a set of pairs $\langle F_i, w_i \rangle$, where F_i is a formula in first-order logic (FOL) and w_i is a real-valued weight. For each finite domain of discourse D, a ground Markov random field (MRF) can be instantiated by introducing to the MRF a Boolean variable for each ground atom and a binary feature $\widehat{f}_j : \mathcal{X} \mapsto \{0, 1\}$ for each ground formula $\widehat{F}_j$, whose value for a possible world $x \in \mathcal{X}$ is 1 if the respective ground formula is satisfied in x and 0 otherwise, and whose weight is w_j. The ground MRF specifies a probability distribution over the set of possible worlds $\mathcal{X}$ according to

$$P(X = x) = \frac{1}{Z} \exp\left(\sum_{j=1}^{|G|} \hat{w}_j \hat{f}_j(x)\right), \tag{1}$$

where Z is a normalization constant and G is an indexed set of weighted ground formulas, i.e. a set of pairs $\langle \widehat{F}_j, \widehat{w}_j \rangle$ containing a pair $\langle \widehat{F}_j, \widehat{w}_j{=}w_i \rangle$ for every ground formula $\widehat{F}_j$ of the formula F_i, and $\widehat{f}_j$ is the feature associated to the j-th pair.

Description Logics The formulas in our probabilistic KBs are not independent of each other. Rather there are many constraints between them. For example, if an entity e is an instance of the concept *Cup* then there might be another entity e' such that the relation *holds*$(\cdot, \cdot)$ holds for the pair $\langle e, e' \rangle$, i.e. the assertion *holds*(e,e') must hold. DL are appropriate representation mechanisms to state such relations. In DL, these constraints are asserted as terminological axioms of the form $c \doteq exp$. In our case, we can assert, for instance, *Cup* $\doteq$ *Container* $\sqcap \exists$ *holds.Liquid* $\sqcap \exists$ *has.Handle* in order to state that the concept of a cup is the intersection of the concept of a container that has a handle and holds some liquid. For the purpose of this work it is important to note that the concept that is defined 'inherits' the constraints from the concepts it is defined with forming a taxonomy relation $\sqsubseteq$. Therefore, the similarity of the relational structure of concepts is highly correlated with their distance in the concept taxonomy. $\top$ denotes the set of all concepts in $\sqsubseteq$.

Semantic similarity The semantic similarity of two concepts in DL-based representations can be characterized in terms of the relative location of the two concepts in the taxonomy. Popular measures take into account the lengths of the shortest paths between two concepts in the respective taxonomy graph. The shorter the paths connecting the two nodes in the graph are, the more similar the respective concepts are assumed to be. Among those similarity measures, the WUP similarity [Wu & Palmer, 1994] $\sim_{\sqsubseteq}: \top \times \top \mapsto [0,1]$ is the most widely used. It defines the semantic similarity on concepts in a class taxonomy as $c_1 \sim_{\sqsubseteq} c_2 := \frac{2 \cdot \mathit{depth}(\mathit{lcs}(c_1,c_2))}{\mathit{depth}(c_1)+\mathit{depth}(c_2)}$, where $lcs(\cdot,\cdot)$ denotes the least common super-concept of two concepts in $\sqsubseteq$.

Fuzzy Logic As we want to reason about concepts that are not contained in our probabilistic model, we need representational means to express our expectations about the properties of an unknown concept, which we are uncertain of. To do this, we intend to replace the binary truth values in MLNs with degrees of beliefs about whether or not relations hold for a concept not contained in the probabilistic model. We use *fuzzy logic* (FL) for this purpose, a multi-valued extension of propositional logic (PL). FL has its foundations in the theory of fuzzy sets, in which elements belong to a set only to a certain degree. Formally, a fuzzy subset x of a set X is a pair $\langle X, \pi_x \rangle$, where X is called the *universe* and $\pi_x : X \mapsto [0,1]$ determines the degree to which a particular element actually belongs to x, which is called the *membership function*. In FL, the universe X is given by the set of atomic propositions and π_x is a fuzzy interpretation of X assigning every proposition in X a real-valued degree of truth. It provides a calculus analogous to the calculus of PL: If A and B are propositions in FL, then the logical connectors with respect to x are defined as $A \wedge B := \min(\pi_x(A), \pi_x(B))$, $A \vee B := \max(\pi_x(A), \pi_x(B))$, and $\neg A := 1 - \pi_x(A)$. Note that the multi-valued logical calculus of FL reduces to its binary counterpart of PL in the extreme cases where all propositions have boolean truth values.

FUZZY-MLNs

A FUZZY-MLN F *is a pair* $\langle L, \sqsubseteq \rangle$*, where* L *is an MLN and* $\sqsubseteq$ *is a taxonomy of concepts, such that* L *represents a conditional probability distribution*

$$P(\mathit{instance\text{-}of}(\cdot,\cdot), \ldots \mid \mathit{is\text{-}a}(\cdot,\cdot), \ldots). \tag{2}$$

In addition, the following conditions hold:

1. *an entity* e *in the domain of discourse* D *is connected to a concept* c *in the taxonomy* $\sqsubseteq$ *always by a proposition* $\mathit{instance\text{-}of}(e,s) \wedge \mathit{is\text{-}a}(s,c)$*, where* $s, c \in \top$*,*
2. *all ground atoms of the form* $\mathit{is\text{-}a}(s,c)$*, where* $s, c \in \top$ *take real-valued degrees of truth* $\in [0,1]$*, which we call* semantic similarity. *Ground atoms of all other predicates take strictly binary truth values* $\in \{0,1\}$*.*
3. *The set* $\mathcal{X}$ *of* possible worlds *represented by* F *is the set of all fuzzy subsets of all ground atoms* X*, where the membership functions for every ground atom is-a(s,c) is equal across all possible worlds and is defined as the semantic similarity of* s *and* c *with respect to* $\sqsubseteq$*, i.e. for all* $x \in \mathcal{X}$ *and for all* $s, c \in \top$*:* $\pi_x(\mathit{is\text{-}a}(s,c)) = s \sim_{\sqsubseteq} c$*.*

In the following, we motivate this definition in more detail.

Probabilistic Semantics According to the second condition in our definition, the semantics of FUZZY-MLNs differs from the original in Equation (1) in two aspects: First, a possible world x is no longer a strictly binary vector assigning a truth value to every ground atom but also allows for real-valued degrees of truth. The ground MRF of a FUZZY-MLN thus contains binary *and* numerical random variables; a real-valued variable for every ground atom of the form *is-a*$(\cdot,\cdot)$ and a binary one for every other ground atom. Second, as a consequence, the semantics of the binary logical features $\hat{f}_j : \mathcal{X} \mapsto \{0,1\}$ in the ground MRF is not applicable any more. We therefore define the features associated to every ground formula $\widehat{F}_j$ in the MRF to take the form $\hat{f}_j : \mathcal{X} \mapsto [0,1]$, where each feature $\hat{f}_j(x)$ evaluates to the truth value of its ground formula $\widehat{F}_j$ in x by applying the fuzzy logic calculus as described above, i.e. $\widehat{f}_j(x) = \pi_x(\widehat{F}_j)$. Hence the distribution of F becomes

$$P(X=x) = \frac{1}{Z} \exp\left(\sum_{j=1}^{|G|} \hat{w}_j \pi_x(\widehat{F}_j) \right). \tag{3}$$

Condition no. 3 in our definition ensures that the probability distribution in (3) corresponds to the conditional distribution in (2): since the truth value of a ground atom of the *is-a* predicate is required to be equal across all possible worlds x, the distribution $P(X{=}x)$ in (3) is effectively conditioned on every atom of the form *is-a*$(\cdot,\cdot)$.

A FUZZY-MLN contains two dedicated predicates, *instance-of* and *is-a*, which provide means to incorporate knowledge from the class taxonomy into the probabilistic model. In short, *is-a* encodes the taxonomic knowledge and *instance-of* is used for expressing uncertainty about which categories entities belong to. By differentiating between the two predicates it can be modelled that one is certain about the taxonomic structure of the domain subject to reasoning but possibly uncertain about which concept an entity belongs to.

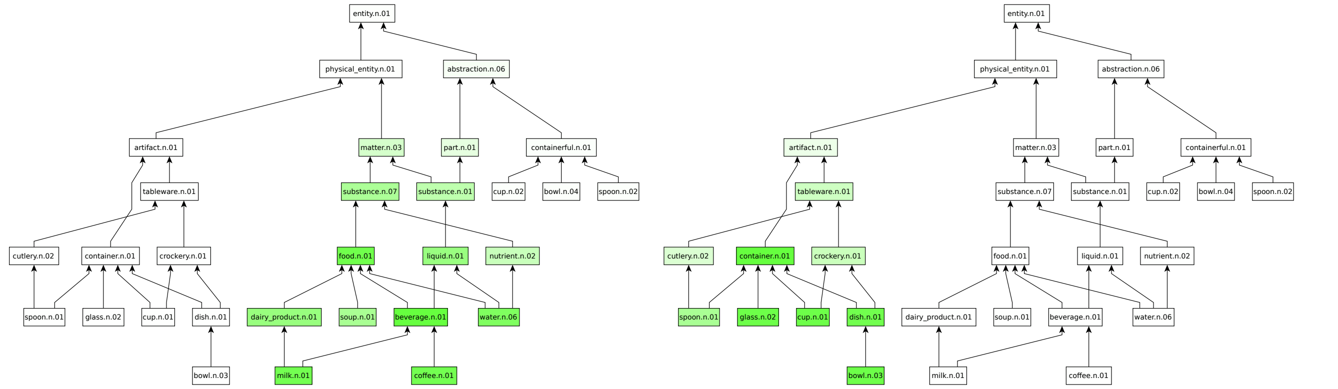


Figure 2: Posteriors distributions over the taxonomy conditioned on semantic roles of a filling action according to Eq. (4). More intense node colors indicate higher probability. *Left:* w_1 given *sem_role*$(w_1, \textit{theme})$ *Right:* w_2 given *sem_role*$(w_2, \textit{goal})$.

In contrast to MLNs, FUZZY-MLNs do not require all predicates to be boolean. Variables (ground atoms) of the form *is-a*(s, c) in the ground MRF take real-valued degrees of truth $\in [0, 1]$, which express the degree to which s is similar to c. Here, a value of 1 denotes maximal similarity, whereas 0 denotes maximal dissimilarity. This allows to represent entities that belong to concepts not contained in the probabilistic knowledge base by referring to them in terms of their similarity to known concepts. Note that in FUZZY-MLNs, the semantic similarities do not have to be computed by probabilistic inference as in other formalisms such as PSL. Rather, they are always given by the taxonomy structure and exclusively appear as evidence. This makes the representation of the conditional distribution in (2) very compact, since the taxonomy structure may be collapsed into single numeric values, which scale the contribution of every single ground formula to the probability mass (3) by the similarities of its constituents to concepts that are contained in the model. This allows to generalize the learned knowledge also to classes unseen in the learning data. In addition, realizing FUZZY-MLNs without having to equip them with the capability of reasoning about the similarity relation $\sim_{\sqsubseteq}$ as such, enables us to escape a complexity monster. Without making this restriction, inference and learning would require us to compute integrals over those variables, rendering computational complexity infeasible for practical applications.

Since the distribution of a FUZZY-MLN is conditioned on the taxonomic structure of the domain, the second predicate, *instance-of*, is used to link any entity in the domain of discourse to a concept in $\sqsubseteq$. Unlike *is-a*, *instance-of* is boolean and may be subject to inference. Propositions about class memberships of an entity e are made in the form *instance-of*$(e, s) \land$ *is-a*(s, c).

Let a minimalistic example illustrate how inference about unknown concepts can be achieved in FUZZY-MLNs: Suppose we want to represent the conditional distributions that parrots can fly and that mammals can not. In Markov logic, we can establish these distributions in an MLN with, for example, the two weighted formulas

$$\begin{aligned} w_1 = \ln(0.9/0.1) \quad & \textit{flies}(e) \land \textit{instance-of}(e, \textit{parrot.n.01}) \\ & \land \textit{is-a}(\textit{parrot.n.01}, \textit{parrot.n.01}) \\ w_2 = \ln(0.1/0.9) \quad & \textit{flies}(e) \land \textit{instance-of}(e, \textit{mammal.n.01}) \\ & \land \textit{is-a}(\textit{mammal.n.01}, \textit{mammal.n.01}). \end{aligned}$$

In classical MLNs, reasoning can only be performed about instances of either of the concepts *parrot.n.01* and *mammal.n.01* because for any other concept, none of the formulas is applicable. Using a FUZZY-MLN with the same model structure and an underlying taxonomy $\sqsubseteq$, however, we can tackle reasoning tasks outside the model domain, such as $P(\textit{flies}(\textit{Fred}) \mid \textit{instance-of}(\textit{Fred}, \textit{turkey.n.01}))$. In this example, there are two ground atoms of the *is-a* predicate, *is-a*(*turkey.n.01*, *parrot.n.01*) and *is-a*(*turkey.n.01*, *mammal.n.01*), which are, for instance, assigned the truth values

$$\begin{aligned} \pi_x(\textit{is-a}(\textit{turkey.n.01}, \textit{parrot.n.01})) = 0.90 \\ \pi_x(\textit{is-a}(\textit{turkey.n.01}, \textit{mammal.n.01})) = 0.01 \end{aligned}$$

in every possible world x according to a similarity $\sim_{\sqsubseteq}$. Consequently, the influence of the two ground formulas

$$\begin{aligned} \widehat{F}_1 = & \textit{flies}(\textit{Fred}) \land \textit{instance-of}(\textit{Fred}, \textit{turkey.n.01}) \\ & \land \textit{is-a}(\textit{turkey.n.01}, \textit{parrot.n.01}) \\ \widehat{F}_2 = & \textit{flies}(\textit{Fred}) \land \textit{instance-of}(\textit{Fred}, \textit{turkey.n.01}) \\ & \land \textit{is-a}(\textit{turkey.n.01}, \textit{mammal.n.01}) \end{aligned}$$

on the distribution in (3) is scaled down by the similarity of concepts. In the extreme case, where there is maximal dissimilarity of two concepts, the contribution of every ground formula vanishes resulting in a uniform distribution. This is reasonable since we cannot make any well-informed statement about entities that are maximally dissimilar to everything that is contained in the model.

Running example continued Let us now continue with our running example and explain how FUZZY-MLNs solve the respective reasoning tasks. We consider again the two training databases corresponding to the instructions (1) ‘fill a glass with milk’ and (2) ‘add a cup of milk’. In order to model word sense and role/sense co-occurrences, we construct a FUZZY-

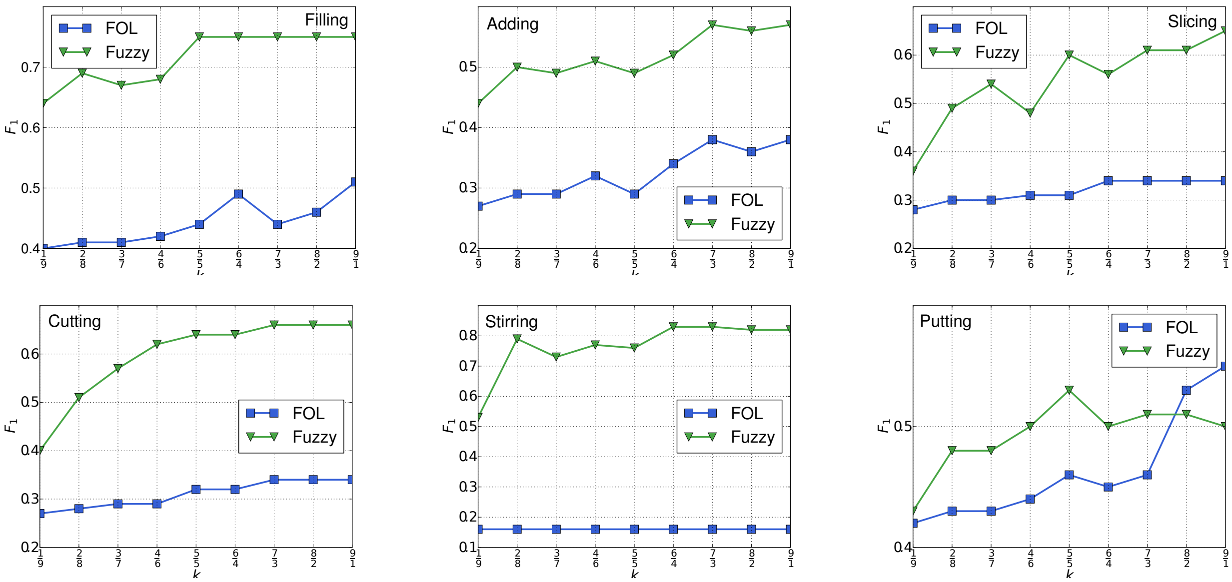


Figure 3: F_1 scores for inverse k-fold cross validation for $k = {}^1\!/\!_9 \ldots {}^9\!/\!_1$ using classical MLNs with FOL semantics and FUZZY-MLNs applied to a WSD problem of 20 examples per action verb.

MLN consisting of one single weighted template formula,

$$\begin{aligned} &\textit{instance-of}(w_1, s_1) \wedge \textit{is-a}(s_1, +c_1) \\ \wedge\ &\textit{instance-of}(w_2, s_2) \wedge \textit{is-a}(s_2, +c_2) \\ \wedge\ &\textit{sem\_role}(w_1, +r_1) \wedge \textit{sem\_role}(w_2, +r_2)\ \wedge\ w_1 \neq w_2, \end{aligned}$$

which has been trained with the two databases introduced at the beginning.

In order to illustrate that the learned MLN indeed reasonably generalizes across classes, we visualize the posterior distributions over the WordNet taxonomy for two exemplary queries. Figure 2 shows the posteriors of two queries for the meaning of a word representing the *theme* of a 'filling' activity and its *goal*, respectively, i.e.

$$P\left(\begin{array}{l} \textit{instance-of}(w_1, s_1), \\ \textit{instance-of}(w_2, s_2) \end{array} \middle| \begin{array}{l} \textit{instance-of}(w', \textit{fill.v.01}), \\ \textit{sem\_role}(w', \textit{action\_verb}), \\ \textit{sem\_role}(w_1, \textit{theme}), \\ \textit{sem\_role}(w_2, \textit{goal}), \\ \textit{is-a}(\textit{fill.v.01}, \textit{fill.v.01}), \ldots \end{array}\right). \tag{4}$$

The distributions show that, conditioned on the semantic role of a word, two clearly separable clusters of concepts loom in the taxonomy. For the *theme* role of a filling action, all substances/liquids gain considerably high probability, whereas the *goal*s of such an action are represented by all types of containers. Note that also categories *not* explicitly modelled, such as *water.n.06*, *soup.n.01*, or *spoon.n.01* and *bowl.n.03* and *glass.n.02*, respectively, have been assigned significant probability masses indicating that the model indeed resonably generalizes across object categories.

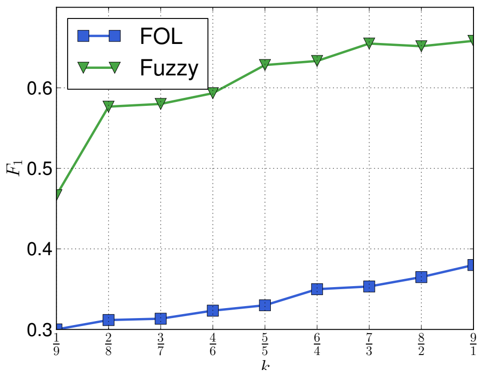
Experiments

We evaluate our method by comparing its performance against classical MLNs with FOL semantics being applied to the problem of word sense disambiguation. We use a real-world data set of natural-language instructions that have been mined from the wikihow.com web site and manually annotated with correct word senses. We take into account sense co-occurrences and part-of-speech tags. The MLN thus only contains one single template formula,

$$\begin{aligned} &\textit{has\_pos}(w_1, +p_1) \wedge \textit{has\_pos}(w_2, +p_2) \\ \wedge\ &\textit{instance-of}(w_1, s_1) \wedge \textit{is-a}(s_1, +c_1) \\ \wedge\ &\textit{instance-of}(w_2, s_2) \wedge \textit{is-a}(s_2, +c_2)\ \wedge\ w_1 \neq w_2. \end{aligned}$$

In order to showcase the generalization capabilities of FUZZY-MLNs, we chose the hardest experimental setup we can imagine: (1) the datasets have been selected to exhibit maximal entropy with respect to the concepts that are contained in the examples, so that they are as dissimilar as possible, and (2) the model was trained with only very small portions of training data. We conduct 'inverse' k-fold cross-validation, a modification of traditional cross-validation, where also inverse proportions of training and test set sizes are considered. For $k = {}^1\!/\!_9$, for example, we use only 10% of the data available for training the model, and the remaining 90% serve for evaluation. Conversely, $k = {}^9\!/\!_1$ corresponds to classical 10-fold cross validation.

We group the instructions with respect to the action verbs they contain and use 20 examples per action verb in each fold. The results are shown in Figure 3 and 4. FUZZY-MLNs clearly outperform the classical MLNs in almost *every* test case. Moreover, FUZZY-MLNs achieve F_1 scores significantly above 0.5 even with very small portions of training data (cmp. 'filling' with only 10% of the data). The F_1 score measures the classification accuracy wrt. word meanings from the taxonomy assigned to each word in the respective NL instruction. It is interesting to note that, while only moderate improvements in classical MLNs are recorded with



Action Verb		k								
		$1/9$	$2/8$	$3/7$	$4/6$	$5/5$	$6/4$	$7/3$	$8/2$	$9/1$
Filling	FOL	0.40	0.41	0.41	0.42	0.44	0.49	0.44	0.46	0.51
	Fuzzy	**0.64**	**0.69**	**0.67**	**0.68**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**
Adding	FOL	0.27	0.29	0.29	0.32	0.29	0.34	0.38	0.36	0.38
	Fuzzy	**0.44**	**0.50**	**0.49**	**0.51**	**0.49**	**0.52**	**0.57**	**0.56**	**0.57**
Slicing	FOL	0.28	0.30	0.30	0.31	0.31	0.34	0.34	0.34	0.34
	Fuzzy	**0.36**	**0.49**	**0.54**	**0.48**	**0.60**	**0.56**	**0.61**	**0.61**	**0.65**
Cutting	FOL	0.27	0.28	0.29	0.29	0.32	0.32	0.34	0.34	0.34
	Fuzzy	**0.40**	**0.51**	**0.57**	**0.62**	**0.64**	**0.64**	**0.66**	**0.66**	**0.66**
Putting	FOL	0.42	0.43	0.43	0.44	0.46	0.45	0.46	**0.53**	**0.55**
	Fuzzy	**0.43**	**0.48**	**0.48**	**0.50**	**0.53**	**0.50**	**0.51**	0.51	0.50
Stirring	FOL	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
	Fuzzy	**0.53**	**0.79**	**0.73**	**0.77**	**0.76**	**0.83**	**0.83**	**0.82**	**0.82**

Figure 4: *Left:* F_1 scores averaged over all action verbs. *Right:* F_1 scores for inverse k-fold cross validation for $k = 1/9 \dots 9/1$.

increasing amounts of training databases, the most significant performance jumps with FUZZY-MLNs can be observed when only sparse training data is used. In these extreme cases, where concepts occur in the test data that are not contained in the training data, classical MLNs (and all other approaches mentioned in the related work) are inapplicable to perform meaningful reasoning but are forced to randomly guess. This shows that fuzzy inference in MLNs can perform adequate reasoning about concepts in the taxonomy that are not explicitly represented in the probability distribution and have not been seen during training.

Related Work

A couple of frameworks have been proposed to incorporate concept taxonomies and similarity in probabilistic models, such as probabilistic description logics [Lukasiewicz, 2008; Niepert, Noessner, & Stuckenschmidt, 2011] (PDL), tractable Markov logic (TML) [Domingos & Webb, 2012] and probabilistic similarity logic (PSL) [Brocheler, Mihalkova, & Getoor, 2010], which differ from FUZZY-MLNs in basically two fundamental ways: (1) FUZZY-MLNs do *not* postulate uncertainty among the taxonomy structure as such, i.e. the structure itself is not subject to reasoning and (2) FUZZY-MLNs do not model the whole taxonomy in the probabilistic model, but only the concepts seen during training. This makes FUZZY-MLNs a more compact reasoning framework. TML is a subset of Markov logic networks. TML introduces the idea of concept taxonomies in MLNs, but in order to perform reasoning about superclasses, the inheritance relationship of concepts is explicitly represented in the model. By employing semantic similarity as evidence, the taxonomy relation is more compactly encoded in FUZZY-MLNs. PSL uses a formalism similar to FUZZY-MLNs. Unlike FUZZY-MLNs, however, the goal of PSL is rather to reason about degree to which a set of entities are similar to each other. Conversely, in FUZZY-MLNs the taxonomy is fixed and serves for filling gaps in the probabilistic KB. Hybrid MLNs (HMLN) [Wang & Domingos, 2008] extend MLNs to reason about continuous variables. They discern features in hard FOL and numeric features that may be expressed as 'soft' (in)equality constraints. Those constraints are typically connected in a multiplicative way, such that, if a logical constraint evaluates to false, then also a connected numeric feature will have no influence on the probability of the respective possible world. Hence representing semantic similarities in HMLNs does not appear straightforward. The concept of soft evidence [Jain & Beetz, 2010] is closely related to the idea of vague evidence, though it has fundamentally different semantics for it still assumes boolean truth values and soft evidences serve as prior probability constraints on ground atoms.

To the best of our knowledge, none of these approaches can deal with entities that are not part of the probabilistic model in any meaningful way. This is a severe limitation, because they are not capable of exhaustively modelling joint probability distributions of realistic domain sizes. Since learning in first-order probabilistic models remains intractable in the general case, inference and generalization across concepts is essential and outstandingly important for probabilistic relational models to be scalable and applicable to real-world problems.

Conclusions

In this work, we have described the design and the implementation of FUZZY-MLNs, an extension of MLNs that allows us to represent probability distributions over open domains compactly – if complete ontologies are available for these domains. The basic idea underlying FUZZY-MLNs is to explicitly represent only the small subset of concepts that is contained in the training databases. After having learned the probability distribution FUZZY-MLNs can reason about concepts that are not contained in the graphical model but in the taxonomy. They do so by exploiting the fact that the relational structure of concepts in the taxonomy is correlated with the relational structures of the explicitly represented concepts weighted by a notion of semantic similarity. FUZZY-MLNs implement this bias by generalizing the *is-a* assertions for off-domain concepts from boolean truth to real-valued degrees of truth. The degree of truth is then computed based on the semantic similarity of the off-domain concept to those concepts contained in the graphical model.

We have shown that FUZZY-MLNs can perform different probabilistic reasoning tasks in a way that matches our intuitions and can outperform probability distributions learned in the ordinary MLN framework both significantly and substantially.

Acknlowlegments

This work has been supported by the EU FP7 projects *RoboHow* (grant no. 288533) and *ACAT* (grant no. 600578).

References

[Bailey, 1997] Bailey, D. 1997. *When push comes to shove: A computational model of the role of motor control in the acquisition of action verbs*. Ph.D. Dissertation, UNIVERSITY of CALIFORNIA.

[Beltagy & Mooney, 2014] Beltagy, I., and Mooney, R. J. 2014. Efficient markov logic inference for natural language semantics. In *Proceedings of the Fourth International Workshop on Statistical Relational AI at AAAI (StarAI-2014)*, 9–14.

[Brocheler, Mihalkova, & Getoor, 2010] Brocheler, M.; Mihalkova, L.; and Getoor, L. 2010. Probabilistic similarity logic. *Conference on Uncertainty in Artificial Intelligence*.

[Domingos & Webb, 2012] Domingos, P., and Webb, W. 2012. A tractable first-order probabilistic logic. In *Proceedings of the Twenty-Sixth National Conference on Artificial Intelligence*.

[Fellbaum, 1998] Fellbaum, C. 1998. *WordNet: an electronic lexical database*. MIT Press USA.

[Getoor *et al.*, 2007] Getoor, L.; Friedman, N.; Koller, D.; Pfeffer, A.; and Taskar, B. 2007. Probabilistic Relational Models. In Getoor, L., and Taskar, B., eds., *Introduction to Statistical Relational Learning*. MIT Press.

[Jain & Beetz, 2010] Jain, D., and Beetz, M. 2010. Soft Evidential Update via Markov Chain Monte Carlo Inference. In *KI 2010: Advances in Artificial Intelligence, 33rd Annual German Conference on AI*, volume 6359 of *Lecture Notes in Computer Science*, 280–290. Springer.

[Lenat, 1995] Lenat, D. 1995. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM* 38(11):33–38.

[Lukasiewicz, 2008] Lukasiewicz, T. 2008. Expressive probabilistic description logics. *Artificial Intelligence* 172(6–7):852 – 883.

[Niepert, Noessner, & Stuckenschmidt, 2011] Niepert, M.; Noessner, J.; and Stuckenschmidt, H. 2011. Log-linear description logics. In *IJCAI*, 2153–2158.

[Nyga, Balint-Benczedi, & Beetz, 2014] Nyga, D.; Balint-Benczedi, F.; and Beetz, M. 2014. PR2 Looking at Things: Ensemble Learning for Unstructured Information Processing with Markov Logic Networks. In *IEEE International Conference on Robotics and Automation (ICRA)*.

[Richardson & Domingos, 2006] Richardson, M., and Domingos, P. 2006. Markov Logic Networks. *Machine Learning* 62(1-2):107–136.

[Sukthankar *et al.*, 2014] Sukthankar, G.; Geib, C.; Bui, H. H.; Pynadath, D.; and Goldman, R. P. 2014. *Plan, Activity, and Intent Recognition: Theory and Practice*. Newnes.

[Tenenbaum *et al.*, 2011] Tenenbaum, J. B.; Kemp, C.; Griffiths, T. L.; and Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. *science* 331(6022):1279–1285.

[Wang & Domingos, 2008] Wang, J., and Domingos, P. 2008. Hybrid Markov Logic Networks. In Fox, D., and Gomes, C. P., eds., *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, 1106–1111. AAAI Press.

[Wu & Palmer, 1994] Wu, Z., and Palmer, M. 1994. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, 133–138. Association for Computational Linguistics Morristown, NJ, USA.