

# Inductive Reasoning about Ontologies Using Conceptual Spaces

**Zied Bouraoui**  
Cardiff University, UK  
BouraouiZ@Cardiff.ac.uk

**Shoab Jameel**  
Cardiff University, UK  
JameelS1@Cardiff.ac.uk

**Steven Schockaert**  
Cardiff University, UK  
SchockaertS1@Cardiff.ac.uk

## Abstract

Structured knowledge about concepts plays an increasingly important role in areas such as information retrieval. The available ontologies and knowledge graphs that encode such conceptual knowledge, however, are inevitably incomplete. This observation has led to a number of methods that aim to automatically complete existing knowledge bases. Unfortunately, most existing approaches rely on black box models, e.g. formulated as global optimization problems, which makes it difficult to support the underlying reasoning process with intuitive explanations. In this paper, we propose a new method for knowledge base completion, which uses interpretable conceptual space representations and an explicit model for inductive inference that is closer to human forms of commonsense reasoning. Moreover, by separating the task of representation learning from inductive reasoning, our method is easier to apply in a wider variety of contexts. Finally, unlike optimization based approaches, our method can naturally be applied in settings where various logical constraints between the extensions of concepts need to be taken into account.

## Introduction

Ontologies encode structured knowledge about the concepts and properties of a given domain. They are typically encoded using description logics, and divided in two parts: a TBox, which contains information about the semantic relationships between concepts, and an ABox, which contains information about which entities belong to what concepts. Similar to ontologies, knowledge graphs encode information about concepts and entities as a set of subject/predicate/object triples. The key difference with ontologies is that knowledge graphs are considerably less expressive (e.g. we cannot express that two concepts are disjoint, or that the intersection of two concepts is subsumed by another concept). To date, ontologies have mostly been used in applications where deductive reasoning suffices, whereas knowledge graphs are often used in combination with forms of inductive reasoning (Lao, Mitchell, and Cohen 2011; Bordes et al. 2013).

As most ontologies are incomplete, it is clearly of interest to combine the deductive machinery of description logics

with some form of induction. This is especially true in applications such as information retrieval, where getting a plausible answer may be better than not getting an answer at all, even if the correctness of that answer cannot be guaranteed. A key requirement in such applications, however, is that answers can be supported by intuitive explanations, as this allows users to determine their plausibility. Another important requirement is that the inductive reasoning process should be relatively cautious, as in a deductive setting the impact of accepting incorrect conclusions could be far-reaching. Finally, conclusions which are derived inductively should be associated with commensurable confidence scores that can be used to restore consistency when conflicts arise. Clearly, approaches which rely on solving a global optimization problem (Bordes et al. 2013) or on aggregating large amounts of individually weak pieces of evidence (Lao, Mitchell, and Cohen 2011) do not meet these requirements.

In this paper, we propose a method which is inspired by cognitive models of category based induction (Osherson et al. 1990; Tenenbaum and Griffiths 2001). In particular, given the knowledge that concepts  $C_1, \dots, C_k$  all have some property  $P$ , our aim is to determine the degree to which we can plausibly conclude that some other concept  $C$  has property  $P$ . To this end, we first rely on the method from (Jameel and Schockaert 2016) to learn a vector space representation of the entities of each considered semantic type (e.g. people, countries, movies), which can be interpreted as a conceptual space (Gärdenfors 2000). From these conceptual spaces, we subsequently derive a set of representative and interpretable features for each semantic type. The plausibility that  $C$  has property  $P$  is then obtained using a form of Bayesian inference over the resulting feature values. We argue that this approach indeed meets the three aforementioned criteria: as it is similar in spirit to cognitive models for category based induction, and relies on interpretable representations, supporting explanations can be derived in a natural way, while the use of Bayesian inference supports cautious forms of interest and yields confidence scores in a principled way. Unlike most existing methods, our method can be used for both TBox and ABox reasoning.

The remainder of this paper is structured as follows. In the next section we give an overview of related work, after which we recall some basic notions from description logics and the theory of conceptual spaces. Subsequently we de-

scribe our proposed model. Finally, we discuss experimental results and present our conclusions.

## Related Work

A natural form of inductive inference consists in using similarity based arguments of the form “if the instance(s) most similar to entity  $x$  belong to concept  $C$  then  $x$  also belongs to concept  $C$ ”. Several variants of this idea, which also underpins nearest neighbor classifiers, have already been explored. In some cases, similarity is obtained from vector space representations that have been learned from text collections (Summers-Stay, Voss, and Cassidy 2016; Beltagy et al. 2013), whereas other approaches derive a notion of similarity directly from the ontology (d’Amato et al. 2010; Minervini et al. 2016; Janowicz 2006). A third class of methods explicitly encodes aspects of similarity as part of the knowledge base (Sheremet et al. 2007).

While intuitive, similarity based reasoning is often too heuristic, especially for TBox reasoning. One particular problem is the context-dependent nature of similarity: it may not be a priori clear in what features two entities need to be similar to draw plausible conclusions about a given concept  $C$ . Interestingly, the method in (Minervini et al. 2016) addresses this by explicitly learning a similarity relation from the relations encoded in the ontology. In particular, they identify relations that are predictive of the fact that two entities have the same value for a given property (e.g. people linked by a friendship relation tend to live in the same country), as well as relations that are predictive of the fact that two entities have a different value (e.g. people linked by the ‘married to’ relation tend to have different values for the property ‘gender’). This form of inference, which relies on the relations between entities, is essentially complementary to the method we propose in this paper.

A second limitation of similarity based reasoning is that it cannot be used when the knowledge base lacks entities that are sufficiently similar to the considered entity  $x$ . The use of analogical proportion based reasoning has been proposed as a technique for alleviating this issue (Miclet, Bayouhdh, and Delhay 2008; Bounhas, Prade, and Richard 2014). Intuitively, for triples of examples  $(a, b, c)$ , a fourth example  $x$  is constructed such that  $a : b :: c : x$  makes an analogical proportion (i.e.  $a$  relates to  $b$  like  $c$  relates to  $x$ ), and similarity based reasoning is applied on the resulting extended set of examples (Hug et al. 2016). Another solution is to use interpolation; properties that hold for  $a$  and  $b$  are then assumed to hold for all entities which are approximately between  $a$  and  $b$  (Schockaert and Prade 2013; Derrac and Schockaert 2015).

The vast majority of existing methods only focus on ABox reasoning, i.e. the problem of identifying which instances (plausibly) belong to which concepts. This problem can be cast as a standard classification problem. One possibility is to make use of kernels for structured data (Bloehdorn and Sure 2007), in combination with classifiers such as support vector machines (SVMs). Intuitively, the classifier then uses the ontology itself to estimate how similar any two given entities are. Another possibility, considered in

(Neelakantan and Chang 2015), is to represent entities using external features derived from resources such as Wikipedia.

Recently, the use of vector space embeddings has become a popular approach for completing knowledge graphs (Bordes et al. 2011; 2013; Wang et al. 2014). These methods focus on finding plausible relationships, intuitively based on statistical correlations with other relationships from the knowledge graph. Some methods of this kind could in principle also be used for inductive ABox reasoning. A number of embedding models that explicitly represent semantic types have recently been proposed (Hu et al. 2015; Xie, Liu, and Sun 2016; Jameel and Schockaert 2016).

One of the main limitations of classification and embedding based methods is that they require sufficient amounts of data to be effective. Moreover, as they are based on black box models, providing intuitive explanations for inferred conclusions can be problematic. In this sense, such methods are complementary to similarity based reasoning and related cognitively inspired methods. The method we propose in this paper aims to combine the best of both worlds.

## Background

Our method is based on two complementary frameworks for representing and reasoning about conceptual knowledge: we use deductive inference in description logics to reason about structured knowledge, and use conceptual space representations as the basis for inductive inferences. In this section, we briefly recall the basic notions from these two frameworks.

### Description Logics

Description logics are the logical framework underlying ontology languages such as OWL. For simplicity, we will consider  $\mathcal{ALC}$  (Baader et al. 2003), which is one of the most basic description logics. However, the results of this paper can straightforwardly be extended to other description logics, as well as related formalisms such as existential rules (Baget et al. 2011; Cali, Gottlob, and Pieris 2012).

The syntax of  $\mathcal{ALC}$  is defined over three pairwise disjoint sets of names  $N_C$ ,  $N_R$  and  $N_I$ , which respectively denote concepts (e.g. *Scientist*), roles (i.e. relations such as *SpecializesIn*) and individuals (i.e. entities such as *einstein*). A concept expression can be (i) an atomic concept from  $N_C$ , (ii) one of the special concepts  $\perp$  and  $\top$ , denoting respectively an empty concept and the universe of all individuals, or (iii) built from other concept expressions using the constructs shown in Table 1. The operators  $\sqcup$ ,  $\sqcap$  and  $\neg$  correspond to the usual notions of union, intersection and complement. The concept expressions  $\exists R.C$  and  $\forall R.C$  depend on a role  $R$  and concept  $C$ . They respectively denote the set of individuals who are related through  $R$  with at least one instance of  $C$ , and the set of individuals who are related through  $R$  only with instances from  $C$ . For example  $\exists \text{SpecializesIn.ScienceArea}$  intuitively denotes the set of individuals who have a specialization that is an area of science.

A knowledge base (KB) is a pair  $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$  where  $\mathcal{T}$  is called the TBox and  $\mathcal{A}$  is called the ABox. The TBox is a set of inclusion axioms of the form  $C \sqsubseteq D$ , where  $C$  and  $D$  are concept expressions. The axiom  $C \sqsubseteq D$  expresses that every instance of  $C$  is also an instance of  $D$ , e.g.

$Scientist \sqsubseteq SpecializesIn.ScienceArea$  expresses that every scientist specializes in some area of science. The ABox contains a set of assertions of the form  $A(a)$  and  $R(a, b)$ , where  $A \in N_C$ ,  $R \in N_R$  and  $a, b \in N_I$ .

An interpretation is a pair  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ , where  $\Delta^{\mathcal{I}}$  is a non-empty set called the domain, and  $\cdot^{\mathcal{I}}$  is a function that assigns to each  $a \in N_I$  an element  $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ , to each  $A \in N_C$  a subset  $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$  and to each  $R \in N_R$  a binary relation  $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ . The function  $\cdot^{\mathcal{I}}$  is extended in a straightforward way for concept and role expressions as shown in Table 1. An interpretation  $\mathcal{I}$  satisfies a concept inclusion axiom, written  $\mathcal{I} \models C \sqsubseteq D$  iff  $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ . Similarly, we say that  $\mathcal{I}$  satisfies a concept (resp. role) assertion, written  $\mathcal{I} \models A(a)$  (resp.  $\mathcal{I} \models R(a, b)$ ), iff  $a^{\mathcal{I}} \in A^{\mathcal{I}}$  (resp.  $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$ ). An interpretation  $\mathcal{I}$  satisfies KB  $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$  if  $\mathcal{I}$  satisfies every the concept inclusion axiom in  $\mathcal{T}$  and every assertions in  $\mathcal{A}$ . If  $\mathcal{I}$  satisfies  $\mathcal{K}$  then  $\mathcal{I}$  is said to be a model of  $\mathcal{K}$ .

Syntax	Semantics
$\top$	$\Delta^{\mathcal{I}}$
$\perp$	$\emptyset$
$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
$\exists R.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} \text{ st } (x, y) \in R^{\mathcal{I}} \text{ and } y \in C^{\mathcal{I}}\}$
$\forall R.C$	$\{x \in \Delta^{\mathcal{I}} \mid \text{if } (x, y) \in R^{\mathcal{I}} \text{ then } y \in C^{\mathcal{I}}\}$

Table 1:  $\mathcal{ALC}$  Syntax and Semantics.

## Conceptual Spaces

Conceptual spaces were proposed by Gärdenfors (2000) as an intermediate representation level between symbolic and connectionist representations. A conceptual space is defined as the Cartesian product of a number of so-called quality dimensions. These can be understood as the (primitive) features in the domain being modelled. For example, in a conceptual space of people, the quality dimensions could correspond to features such as height, age, education level, etc. Each entity of the considered domain is thus represented by a vector  $(x_1, \dots, x_n)$ , where  $x_i$  is the value of this entity for the  $i^{\text{th}}$  quality dimension. In general, quality dimensions can be arbitrary metric spaces in which a notion of betweenness is defined. However, in this paper, we will make the common simplifying assumption that each quality dimension is modelled as the real line.

## Inductive Reasoning Using Embeddings

To obtain conceptual space representations, we have used the method from (Jameel and Schockaert 2016), which uses textual descriptions from Wikipedia and structured information from Wikidata as input. For each semantic type occurring in Wikidata, this method constructs a Euclidean space in which all entities of that type are represented as points. The number of dimensions of each space is chosen in an automated way, depending on how much information is available about the corresponding semantic type. Moreover, the words occurring in the textual descriptions of these entities are represented as vectors in this Euclidean space. The idea

is that each of these vectors could potentially be defining a quality dimension. When learning the space, the following constraint is imposed for each entity  $i$  and context word  $j$ :

$$p_i v_j \approx \log c_{ij} - b_i - b_j \quad (1)$$

where  $p_i$  is the vector representing entity  $i$ ,  $v_j$  is the vector representing the context word  $j$ ,  $c_{ij}$  is the number of times word  $j$  occurs in the description of entity  $i$ , and  $b_i, b_j \in \mathbb{R}$  are constants. Since the resulting optimization problem is over-constrained, (1) will only be (approximately) satisfied for some words, and these words tend to be those that correspond to salient properties. Therefore, to select quality dimensions, we determine those words  $j$  for which (1) is maximally satisfied. Specifically, for each word  $j$ , we define:

$$e_j = \sum_i (p_i v_j - \log c_{ij} + b_i + b_j)^2$$

To allow for some redundancy, for semantic types that are modelled in  $\mathbb{R}^n$  we will select  $2n$  quality dimensions. One possibility would be to select the  $2n$  terms for which  $e_j$  is minimal. This choice, however, is likely to be suboptimal if among these terms we find near-synonyms. Therefore, we instead select the  $5n$  terms that minimize  $e_j$  and then cluster the corresponding vectors in  $2n$  clusters, using  $k$ -means clustering with cosine similarity. The mean vector  $v_C$  for each cluster  $C$  is then treated as a quality dimension. The coordinate of an entity  $i$  for that quality dimension is defined as  $p_i \cdot v_C$ . Moreover, note that each quality dimension is associated with a small set of labels, which enables us to generate natural language explanations for inductive inferences.

## Inductive ABox Reasoning

Let  $B$  be a direct subconcept of  $A$ , i.e.  $B \sqsubseteq A$  is in the TBox of the considered ontology and there is no other concept  $C$  such that  $B \sqsubseteq C$  and  $C \sqsubseteq A$  are both in the TBox. Suppose that  $b_1, \dots, b_m$  are the known instances of  $B$ , and that  $a_1, \dots, a_n, b_1, \dots, b_m$  are the known instances of  $A$ . The problem we consider is to determine which elements from  $\{a_1, \dots, a_n\}$  are likely to be instances of  $B$ . Note that in practice, we would typically have a concept hierarchy, where  $B_1 \sqsubseteq B_2 \sqsubseteq \dots \sqsubseteq B_{k-1} \sqsubseteq B_k \sqsubseteq \top$ , and first determine instances that are likely to be in  $B_k$ ; among those instances, we can then determine the ones that are likely to be in  $B_{k-1}$  (and similar for the other subconcepts of  $B_k$ ), etc.

The first step of our method consists in linking the individuals  $a_1, \dots, a_n, b_1, \dots, b_m$  to Wikidata entities, and thus to a conceptual space representation. If, for the considered ontology, there exists an explicit mapping to Wikidata, or to a resource that is linked to Wikidata such as WordNet, this step is trivial. In other cases, we make use of BabelNet (Navigli and Ponzetto 2012) to suggest likely matches. To deal with ambiguities in the latter case, we first remove candidate links whose semantic type in Wikidata does not match that of the majority of the other instances of  $A$ . If any ambiguity remains, we can either disregard that instance for inductive reasoning, or default to the most common sense (e.g. choosing the Wikidata entity whose associated Wikipedia page is

largest). Let  $\tau_1, \dots, \tau_k$  be the Wikidata types that have at least 80% of the entities  $a_1, \dots, a_n, b_1, \dots, b_m$  as instances. Note that there are typically several such types, because the types are organised in a directed acyclic graph through the *subclass* relationship, and because entities may have types that relate to different aspects (e.g. *Eiffel Tower* is an instance of both *tower* and *landmark*). The following inductive inference process is repeated for each of the types  $\tau_1, \dots, \tau_k$ , and the plausible assertion  $B(a_j)$  is added to the ABox as soon as the conceptual space associated with one of these types warrants it, as we describe next.

Let  $v_1, \dots, v_k$  be the vectors defining the quality dimensions of the conceptual space for some type  $\tau$ . We denote the coordinate for the  $i^{\text{th}}$  quality dimension as  $x_i^j = a_j v_i$  for the instances  $a_j$  and as  $y_i^j = b_j v_i$  for the instances  $b_j$ . The probability that  $a_j$  belongs to concept  $B$ , given that we already know that it belongs to concept  $A$  and that its coordinate is  $x_i^j$ , can be written as:

$$p(B | x_i^j, A) = \frac{p(x_i^j | B, A) \cdot p(B | A)}{p(x_i^j | A)}$$

A central assumption we make is that for each quality dimension  $i$  the coordinates of instances of each concept  $A$  follow a Gaussian distribution  $G_{(A,i)}$ . Recall that the product of two Gaussians is another Gaussian (up to renormalization), hence we can think of  $G_{(B,i)}$  as the product of  $G_{(A,i)}$  and some other Gaussian, which is in line with the view that the instances of  $B$  are those instances of  $A$  that satisfy some particular properties. Given that  $p(B | A)$  is unknown, all we can say is that the probability that  $a_j$  belongs to  $B$  is proportional to  $\prod_i \frac{p(x_i^j | B, A)}{p(x_i^j | A)} = \prod_i \frac{G_{(B,i)}(x_i^j)}{G_{(A,i)}(x_i^j)}$ , where we additionally rely on the assumption that all quality dimensions provide independent evidence. This leads us to use the following log-likelihood ratio as a scoring function:

$$\text{conf}(a_j, B) = \sum_{i=1}^k \log G_{(B,i)}(x_i^j) - \log G_{(A,i)}(x_i^j) \quad (2)$$

Since the parameters of the Gaussians  $G_{(A,i)}$  and  $G_{(B,i)}$  are unknown, they have to be estimated from the values  $x_i^j$  and  $y_i^j$ . To this end we use a normal-inverse- $\chi^2$  prior. In particular, we estimate  $G_{(B,i)}$  as follows ( $x \in \mathbb{R}$ ):

$$G_{(B,i)}(x) = \int G(x; \mu, \sigma^2) I_{\chi^2}(\mu, \sigma^2) d\mu d\sigma \quad (3)$$

where  $G(\cdot; \mu, \sigma^2)$  denotes a Gaussian distribution, and  $I_{\chi^2}$  is an inverse- $\chi^2$  distribution with parameters  $\mu_{(B,i)}, \sigma_{(B,i)}^2, \nu_{(B,i)}, \kappa_{(B,i)}$ . In absence of prior knowledge, these parameters can be estimated from the data points  $y_i^1, \dots, y_i^m$  as follows (Murphy 2007):

$$\begin{aligned} \mu_{(B,i)} &= \frac{1}{m} \sum_j y_i^j & \sigma_{(B,i)}^2 &= \frac{1}{m-1} \sum_j (y_i^j - \mu_{(B,i)})^2 \\ \nu_{(B,i)} &= m-1 & \kappa_{(B,i)} &= m \end{aligned}$$

The use of Bayesian inference in cognitive models of induction has been advocated, among others in (Tenenbaum

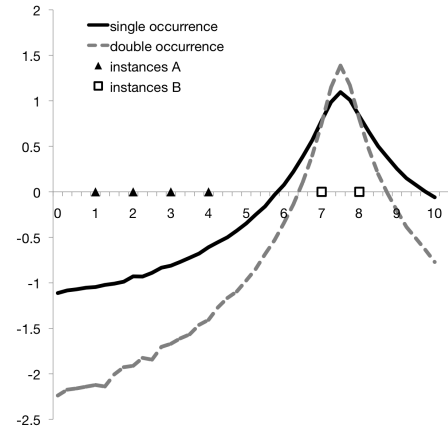


Figure 1: Impact of data sparsity.

and Griffiths 2001). In contrast to the model proposed in (Tenenbaum and Griffiths 2001), however, we estimate inductive strength using a likelihood ratio. This is important in our context because we cannot be certain that the available quality dimensions are sufficient to distinguish between instances from  $A$  and  $B$ . If the available representation is indeed insufficient, then we would expect  $G_{(B,i)}$  to be similar to  $G_{(A,i)}$  for each  $i$  and thus  $\text{conf}(a_j, B) \approx 0$ . In contrast, if  $\text{conf}(a_j, B) \gg 0$  we know that the quality dimensions are sufficient to conclude that  $a_j$  is likely to be an instance of  $B$ . Note that if we assume a uniform prior  $P(B|A) = 0.5$ , then  $\text{conf}(a_j, B) > 0$  is equivalent to  $p(B | x_i^j, A) > 0.5$ .

Compared to maximum likelihood estimates, the use of Bayesian estimates in (3) leads to more cautious predictions when the number of data points is small, which is important because we only want assertions to be added to the ontology if the available evidence is sufficiently strong. Figure 1 illustrates this effect: the solid line shows the value of  $\text{conf}(\cdot, B)$  in a scenario where there is only one quality dimension, the instances in  $B$  have coordinates 7, 8 and the remaining instances of  $A$  have coordinates 1, 2, 3, 4. The dashed line shows a similar scenario, with the only difference that each coordinate appears twice, i.e. the instances in  $B$  have coordinates 7, 7, 8, 8 and the remaining instances of  $A$  have coordinates 1, 1, 2, 2, 3, 3, 4, 4. As can be seen, in the latter case we, the presence of more data results in more extreme confidence scores.

Note that the labels for the quality dimensions can be used to produce intuitive explanations. To this end, we first determine the quality dimensions  $i$  which maximally impact the confidence degree. For example, if the coordinate  $x_i^j$  of entity  $a_j$  is such that  $G_{(B,i)} \gg G_{(A,i)}$  and the mean of  $G_{(B,i)}$  is higher than the mean of  $G_{(A,i)}$ , an explanation could take the form of “ $a_j$  is believed to be an instance of  $B$ , as it is more (strongly related to)  $L_i$  than typical instances of  $A$ ”, where  $L_i$  is the label for the  $i^{\text{th}}$  quality dimension.

Now consider a setting where some concept  $D$  has three direct subconcepts  $A, B$  and  $C$ . When we use (2) to determine which instances of  $D$  should also be instances of  $A, B$  and  $C$ , conflicts may arise. Consider the scenario illustrated

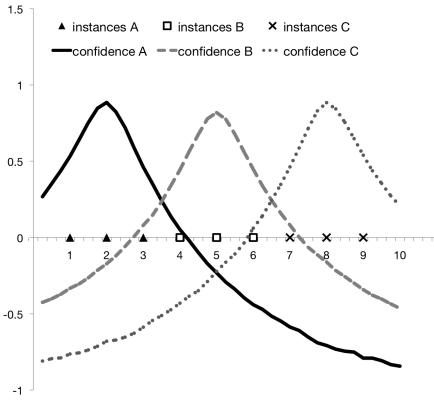


Figure 2: Determining concept boundaries for disjoint categories.

in Figure 2, where we again consider a single quality dimension for simplicity, and the known instances of  $A$  have coordinates 1, 2, 3, the known instances of  $B$  have coordinates 4, 5, 6, the known instances of  $C$  have coordinates 7, 8, 9, and we assume that  $D$  has no other instances. Figure 2 displays the values of  $\text{conf}(\cdot, A)$ ,  $\text{conf}(\cdot, B)$  and  $\text{conf}(\cdot, C)$ . If we conclude that an entity with coordinate  $z$  belongs to  $A$  as soon as  $\text{conf}(\cdot, A) > 0$ , and similar for  $B$  and  $C$ , some entities will be added to more than one subconcept. Hence, if the ontology encodes that these subconcepts are disjoint (e.g.  $A \cap B \subseteq \perp$ ), a conflict arises. A natural way to repair this conflict consists in removing the conclusions with the lowest confidence degree (if this contributes to resolving the conflict). The result is that an entity with coordinate  $z$  is added to  $A$  if  $G_{(B,i)}(z) > \max(G_{(B,i)}(z), G_{(C,i)}(z))$ , and similar for  $B$  and  $C$ . In this way, the inductive and deductive inference process are effectively combined to derive the most appropriate (hard) concept boundaries.

### Inductive TBox Reasoning

We now consider a scenario involving three concepts  $A, B, C$  such that  $B$  and  $C$  are both direct subconcepts of  $A$ . The problem we consider consists in determining to what extent we can be confident that  $C \sqsubseteq B$  is a valid inclusion axiom. As for ABox reasoning, this inductive inference step can be applied in a recursive fashion, resulting in a general method for adding plausible inclusion axioms. Let  $b_1, \dots, b_m$  be the known instances of  $B$ ,  $c_1, \dots, c_n$  the known instances of  $C$  and  $a_1, \dots, a_p, b_1, \dots, b_m, c_1, \dots, c_n$  the known instances of  $A$ . These instances are linked to Wikidata entities, as before, and we again determine the Wikidata semantic types that cover (most) of the instances of  $A$ . To determine whether  $C \sqsubseteq B$  is likely to be valid, we propose the following confidence score:

$$\text{conf}(B, C) = \sum_{i=1}^k \log \int \frac{G_{(B,i)}(x)}{G_{(A,i)}(x)} \cdot G_{(C,i)}(x) dx \quad (4)$$

where  $G_{(A,i)}(x)$ ,  $G_{(B,i)}(x)$  and  $G_{(C,i)}(x)$  are evaluated as in (3). Similar as for ABox reasoning, we assume that there are (unknown) Gaussians associated with the concepts  $A, B$

Table 2: Results for ABox reasoning.

	SR			BR		
	Rec	Pr	F1	Rec	Pr	F1
$ X  < 5$	1.0	1.0	1.0	1.0	1.0	1.0
$ X  < 10$	0.909	0.842	0.874	0.958	0.904	0.930
$ X  < 50$	0.725	0.827	0.773	0.821	0.849	0.835
$ X  \geq 50$	0.794	0.249	0.379	0.706	0.816	0.757

and  $C$ . Rather than evaluating the likelihood ratio (2) for a specific value, in (4) we evaluate what this likelihood ratio is on average for instances of concept  $C$ .

We also consider a second way to use the conceptual space representations for TBox reasoning, which relies on the fact that concepts from the ontology may appear as instances in Wikidata. For example, *scientist* is regarded as an instance of the semantic type *profession*<sup>1</sup>. As a result, in some cases we can determine confidence degrees for inclusion axioms using (2). In particular, let  $B_1, \dots, B_m$  be the subconcepts of  $B$  (according to the ontology) for which we have a conceptual space representation (as an instance), and let  $A_1, \dots, A_n, B_1, \dots, B_m$  be the subconcepts of  $A$  for which this is the case. Then (2) determines the confidence that the inclusion axiom  $A_j \sqsubseteq B$  is valid, where the coordinates  $x_i^j$  are now obtained from the representation of concept  $A_j$  instead of an individual, and similar for the coordinates  $y_i^j$  corresponding to the subconcepts of  $B$ .

### Evaluation

We have implemented<sup>2</sup> our method, using Java including OWLAPI<sup>3</sup> and Pellet reasoner<sup>4</sup> for deductive reasoning tasks (e.g. determining the instances or subconcepts of a given concept, checking consistency), and using a standard approach to repairing inconsistencies which uses the confidence scores as a penalty. To evaluate the effectiveness of the proposed approach, we have used the OWL version of the SUMO ontology<sup>5</sup>. This is a relatively large open-domain ontology, covering a total of 4558 concepts, 86457 individuals, 5330 inclusion axioms and 167381 ABox assertions. An advantage of SUMO is that for several concepts, an explicit mapping to WordNet is provided (which has, in turn, been linked to Wikidata).

We have evaluated our method (which is referred to as BR below) against the following baselines:

**Similarity based reasoning (SR)** For ABox reasoning, to decide whether an instance  $a$  of  $A$  is also an instance of the subconcept  $B$ , we check whether the instance from  $A$  which is most similar to  $a$  belongs to  $B$ . For TBox reasoning, we apply the same method by using the representations of SUMO concepts as instances in Wikidata.

**WordNet (WN)** This method is only used for TBox reasoning. To decide whether a concept  $B$  is subsumed by  $A$ , we

<sup>1</sup><https://www.wikidata.org/wiki/Q901>

<sup>2</sup><http://www.cs.cf.ac.uk/semanticspaces/>

<sup>3</sup><http://owlapi.sourceforge.net>

<sup>4</sup><https://github.com/stardog-union/pellet>

<sup>5</sup><http://www.adampease.org/OP/>

Table 3: Results for TBox reasoning

	Variant 1									Variant 2		
	WN			SR			BR			BR		
	Rec	Pr	F1	Rec	Pr	F1	Rec	Pr	F1	Rec	Pr	F1
$ X  < 5$	0.79	1.0	0.883	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$ X  < 10$	0.632	1.0	0.775	0.914	0.847	0.879	0.938	0.966	0.952	1.0	1.0	1.0
$ X  < 50$	0.214	1.0	0.353	0.490	0.836	0.618	0.806	0.822	0.814	0.954	0.916	0.935
$ X  \geq 50$	0.159	1.0	0.274	0.399	0.639	0.491	0.537	0.714	0.613	0.769	0.823	0.795

check whether the WordNet equivalent of  $B$  is a hyponym of the WordNet equivalent of  $A$ .

We also experimented with the use of Support Vector Machines, but found them to be uncompetitive in this setting, due to the small number of instances in most cases and the lack of true negative examples (i.e. either a one-class set-up needs to be used, or we have to assume that instances which are not known to be in  $B$  are not in  $B$ ).

To generate test instances for ABox reasoning, we consider each concept  $A$  with direct subconcept  $B$ , such that there exists some Wikidata type  $S$  such that  $A$  contains at least four instances of  $S$  and  $B$  contains at least three<sup>6</sup> instances of  $S$ . For the evaluation of ABox reasoning, we use three-fold cross validation<sup>7</sup> as follows. Let  $Y$  be the entities that are asserted to be instances of  $B$  in SUMO, and let  $X$  be the entities that are asserted to be instances of  $A$  (where  $Y \subseteq X$ ). We split  $Y = Y_1 \cup Y_2 \cup Y_3$  in three sets of (approximately) the same size. Given the information that  $X$  are instances of  $A$  and  $Y_2 \cup Y_3$  are instances of  $B$ , each method is then applied to determine which among the instances in  $X \setminus (Y_2 \cup Y_3)$  are also instances of  $B$ . Let  $Z$  be the instances predicted to belong to  $B$ . The performance is evaluated using precision  $Pr$ , recall  $Rec$  and F1, defined as (assuming  $|Z| > 0$ ):

$$Pr = \frac{|Z \cap Y_1|}{|Z|} \quad Rec = \frac{|Z \cap Y_1|}{|Y_1|} \quad F1 = \frac{2 \cdot Pr \cdot Rec}{Pr + Rec}$$

The whole process is then repeated two times, letting respectively  $Y_2$  and  $Y_3$  play the role of  $Y_1$ .

For TBox reasoning, we consider two variants as in the previous section. In the first variant, we use the fact that concepts in SUMO are instances in WikiData, and use the same process as for ABox reasoning. This method can be applied in cases where the concepts  $A$ ,  $B$  and  $C$  each have at least three subconcepts. For our method, we then use (2) to determine confidence. In the second variant, we use our method based on the instances of the concepts  $A$ ,  $B$  and  $C$ , using (4) to determine confidence. This method can be applied in cases where the concepts  $A$ ,  $B$  and  $C$  each have at least three instances. Note that both variants are applicable in different contexts, i.e. variant 1 can be used even if the ABox is empty, but requires the concept  $A$  to have a sufficient number of subconcepts, while variant 2 requires a sufficient number of instances in the ABox. In all cases, when

<sup>6</sup>This restriction is needed to apply three-fold cross-validation.

<sup>7</sup>While 5 or 10 folds are more commonly used, we are especially interested in concepts with very few instances.

using our method, we make the inductive inference as soon as the confidence score is strictly positive.

The results for ABox reasoning are shown in Table 2, grouped by the number of instances in the set of individuals  $X$  that are known to belong to the concept  $A$ . Overall we can see that our method outperforms similarity based reasoning, especially in terms of precision for concepts with many instances. In particular, if the majority of instances of  $A$  are not instances of  $B$ , there are very limited guarantees that the most similar instances to the entity being categorized will be instances of  $B$ , which is a result of the context-dependent nature of similarity. Note that, for our method, by setting the threshold for the confidence degree higher than 0, we can increase precision although at the cost of lower recall.

The results for both variants of TBox reasoning are shown in Table 3. Note that the results for both variants are not directly comparable, as the set of concepts to which they were applied differs. For Variant 1, our method outperforms the other methods in terms of F1 score, although the use of WordNet naturally leads to higher precision. The results for Variant 2 are particularly promising, although we are not aware of any baselines against which this method can be directly compared. When this method can be applied, i.e. when sufficient instances are available, it is clear that the richer concept representation it relies on, compared to methods that rely on point representations of concepts (i.e. Variant 1), leads to clear advantages.

## Conclusions

We have proposed a new method for inductive reasoning with ontologies. The method consists in using a form of Bayesian inference over interpretable feature representations that are obtained from a learned vector space embedding. Our experimental results show that this method outperforms similarity based reasoning for both ABox and TBox reasoning, while retaining the ability to generate intuitive explanations.

## Acknowledgments

This work was supported by ERC Starting Grant 637277.

## References

Baader, F.; Calvanese, D.; McGuinness, D. L.; Nardi, D.; and Patel-Schneider, P. F., eds. 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. New York, NY, USA: Cambridge University Press.

- Baget, J.; Leclère, M.; Mugnier, M.; and Salvat, E. 2011. On rules with existential variables: Walking the decidability line. *Artif. Intell.* 175(9-10):1620–1654.
- Beltagy, I.; Chau, C.; Boleda, G.; Garrette, D.; Erk, K.; and Mooney, R. 2013. Montague meets Markov: Deep semantics with probabilistic logical form. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics*, 11–21.
- Bloehdorn, S., and Sure, Y. 2007. Kernel methods for mining instance data in ontologies. In *Proceedings of the 6th International Semantic Web Conference*, 58–71.
- Bordes, A.; Weston, J.; Collobert, R.; and Bengio, Y. 2011. Learning structured embeddings of knowledge bases. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*.
- Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of the Annual Conference on Neural Information Processing Systems*. 2787–2795.
- Bounhas, M.; Prade, H.; and Richard, G. 2014. Analogical classification: A new way to deal with examples. In *Proceedings of the 21st European Conference on Artificial Intelligence*, 135–140.
- Cali, A.; Gottlob, G.; and Pieris, A. 2012. Towards more expressive ontology languages: The query answering problem. *Artif. Intell.* 193:87–128.
- d’Amato, C.; Fanizzi, N.; Fazzinga, B.; Gottlob, G.; and Lukasiewicz, T. 2010. Combining semantic web search with the power of inductive reasoning. In *Proceedings of the 4th International Conference on Scalable Uncertainty Management*. 137–150.
- Derrac, J., and Schockaert, S. 2015. Inducing semantic relations from conceptual spaces: a data-driven approach to plausible reasoning. *Artificial Intelligence* 74–105.
- Gärdenfors, P. 2000. *Conceptual Spaces: The Geometry of Thought*. MIT Press.
- Hu, Z.; Huang, P.; Deng, Y.; Gao, Y.; and Xing, E. P. 2015. Entity hierarchy embedding. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, 1292–1300.
- Hug, N.; Prade, H.; Richard, G.; and Serrurier, M. 2016. Analogical classifiers: A theoretical perspective. In *Proceedings of the 22nd European Conference on Artificial Intelligence*, 689–697.
- Jameel, S., and Schockaert, S. 2016. Entity embeddings with conceptual subspaces as a basis for plausible reasoning. In *Proceedings of the 22nd European Conference on Artificial Intelligence*, 1353–1361.
- Janowicz, K. 2006. Sim-DL: Towards a semantic similarity measurement theory for the description logic ALCNR in geographic information retrieval. In *OTM Confederated International Conferences*, 1681–1692.
- Lao, N.; Mitchell, T.; and Cohen, W. W. 2011. Random walk inference and learning in a large scale knowledge base. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 529–539.
- Miclet, L.; Bayouh, S.; and Delhay, A. 2008. Analogical dissimilarity: Definition, algorithms and two experiments in machine learning. *Journal of Artificial Intelligence Research* 32:793–824.
- Minervini, P.; d’Amato, C.; Fanizzi, N.; and Tresp, V. 2016. Discovering similarity and dissimilarity relations for knowledge propagation in web ontologies. *Journal on Data Semantics* 1–20.
- Murphy, K. 2007. Conjugate Bayesian analysis of the Gaussian distribution. Technical report, University of British Columbia.
- Navigli, R., and Ponzetto, S. P. 2012. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence* 193:217–250.
- Neelakantan, A., and Chang, M. 2015. Inferring missing entity type instances for knowledge base completion: New dataset and methods. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 515–525.
- Osherson, D. N.; Smith, E. E.; Wilkie, O.; Lopez, A.; and Shafir, E. 1990. Category-based induction. *Psychological review* 97(2):185–200.
- Schockaert, S., and Prade, H. 2013. Interpolative and extrapolative reasoning in propositional theories using qualitative knowledge about conceptual spaces. *Artificial Intelligence* 202:86–131.
- Sheremet, M.; Tishkovsky, D.; Wolter, F.; and Zakharyashev, M. 2007. A logic for concepts and similarity. *Journal of Logic and Computation* 17(3):415–452.
- Summers-Stay, D.; Voss, C.; and Cassidy, T. 2016. Using a distributional semantic vector space with a knowledge base for reasoning in uncertain conditions. *Biologically Inspired Cognitive Architectures* 16:34 – 44.
- Tenenbaum, J. B., and Griffiths, T. L. 2001. Generalization, similarity, and bayesian inference. *Behavioral and Brain Sciences* 24:629–640.
- Wang, Z.; Zhang, J.; Feng, J.; and Chen, Z. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 1112–1119.
- Xie, R.; Liu, Z.; and Sun, M. 2016. Representation learning of knowledge graphs with hierarchical types. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2965–2971.