Predicting the quality of semantic relations by applying Machine Learning Classifiers to the Semantic Web

Miriam Fernandez, Marta Sabou, Petr Knoth, Enrico Motta
Knowledge Media Institute (KMi)
The Open University
Walton Hall, Milton Keynes, Mk7 6AA, United Kingdom
{m.fernandez, r.m.sabou, p.knoth, e.motta}@open.ac.uk

ABSTRACT
In this paper, we propose the application of Machine Learning (ML) methods to the Semantic Web (SW) as a mechanism to predict the correctness of semantic relations. For this purpose, we have acquired a learning dataset from the SW and we have performed an extensive experimental evaluation covering more than 1,800 relations of various types. We have obtained encouraging results, reaching a maximum of 74.2% of correctly classified semantic relations for classifiers able to validate the correctness of multiple types of semantic relations (*generic classifiers*) and up to 98% for classifiers focused on evaluating the correctness of one particular semantic relation (*specialized classifiers*).

Categories and Subject Descriptors
I.5.2 [**Pattern Recognition**]: Design Methodology –*Classifier design and evaluation, Feature evaluation and selection, Pattern analysis.*

General Terms
Algorithms, Measurement, Design, Experimentation.

Keywords
Semantic Web, Semantic Relations, Machine Learning.

1. INTRODUCTION
The problem of relation extraction between two terms is a well-known research problem traditionally addressed by the Natural Language Processing (NLP) community. The approaches found in the literature follow several different trends like: the exploitation of lexical patters to extract relations from textual corpora [3], the generation of statistical measures that detect correlations between words based on their frequency within documents [2] or, the exploitation of structured knowledge resources like WordNet[1] to detect or refine relations [1].

With the evolution of the SW notion of knowledge reuse, from an ontology-centered view, to a more fine-grained perspective where individual knowledge statements (i.e., semantic relations) are reused rather than entire ontologies, a parallel problem arises: *estimating the correctness of a known relation between two terms*. As an illustrative example, imagine the two following relations: *Book – containsChapter –Chapter, Chapter ⊆ Book.* While the relation *Book – containsChapter –Chapter* can be considered correct independently of an interpretation context, in the case of *Chapter ⊆ Book,* subsumption has been used incorrectly to model a meronymy relation.

One of the first attempts to address this problem is the work of Sabou et al. [4]. In this wok the authors investigate the use of the Semantic Web (SW) as a source of evidence for predicting the correctness of a semantic relation. They show that the SW is not just a motivation to investigate the problem, but a large collection of knowledge-rich results that can be exploited to address it. Following this idea, the work presented in this paper *makes use of the SW as a source of evidence for predicting the correctness of semantic relations*. However, as opposed to [4], which introduces several evaluation measures based on the adaptation of existing Natural Language methodologies to SW data, this work aims to approach the problem *using Machine Learning (ML) techniques*. For this purpose, we have worked on: a) acquiring a medium-scale learning dataset from the SW and b) performing an experimental evaluation covering more than 1,800 relations of various types. We have obtained encouraging results, reaching a maximum of 74.2% of correctly classified semantic relations for classifiers able to validate the correctness of multiple types of semantic relations (*generic classifiers*) and up to 98% for classifiers focused on evaluating the correctness of one particular semantic relation (*specialized classifiers*).

2. ACQUIRING A LEARNING DATASET
The problem addressed in this work can be formalized as a classification task. In this type of Machine Learning problems, the learning method is presented with a set of classified examples from which it is expected to learn how to predict the classification of unseen examples. The collection of classified examples, or the *learning dataset*, is obtained in three phases. In the **first phase,** a set of manually evaluated semantic relations is acquired. These relations can be seen as a quadruple <*s, R, t, e*> where *s* is the source term, *t* is the target term, *R* is the relation to be evaluated, and *e* {T, F} is a manual Boolean evaluation provided by users where T denotes a *true* or correct relation, and F denotes a *false* or incorrect relation; e.g., <*Helicopter,* ⊆ *, Aircraft, T*>. This experimental data is obtained from the datasets of the Ontology Alignment Evaluation Initiative[2] (OAEI) and includes the AGROVOC/NALT and the OAEI'08 datasets. These datasets comprise a total of 1,805 semantic relations of different types: ⊆, ⊇, ⊥ and named. Among them, 1,129 are evaluated as true (T),

[1] http://wordnet.princeton.edu/

[2] http://oaei.ontologymatching.org/

correct relations, and 676 are evaluated as false (F), incorrect relations. In the **second phase**, a set of SW mappings (occurrences of relations containing the same or equivalent source, *s* and target, *t* terms in the publicly available SW data) is obtained for each particular semantic relation. These mappings are extracted using the services of the Watson SW gateway. Specific details about the SW mapping extraction algorithm can be found in [4]. In the **third phase**, these mappings are formalized and represented in terms of the values of their features (or attributes). The selected attributes to represent each classified example are:

- *e*, the relation correctness {T, F}. This is the *class attribute*, i.e., the one that will be predicted for future examples.
- Type(R), the type of relation to be evaluated: ⊆, ⊇, ⊥ and named relations.
- $|M|$, the number of mappings.
- $|M \subseteq|$, the number of subclass mappings.
- $|M \supseteq|$, the number of superclass mappings.
- $|M \perp|$, the number of disjoint mappings.
- $|M_R|$, the number of named related mappings.
- $|M_S|$, the number of sibbling mappings.
- For each particular mapping M_i we consider
  - Type (R_i), the relation type of the mapping: ⊆, ⊇, ⊥, named and sibling.
  - Pl (M_i) the path length of the mapping M_i
  - Np (M_i) the number of paths that lead to the mapping M_i. Note that for sibling and named mappings the connection can be derived from 2 different paths connected by a common node.
  - $|M_i \subseteq|$, the number of subclass relations in M_i
  - $|M_i \supseteq|$, the number of superclass relations in M_i
  - $|M_i \perp|$, the number of disjoint relations in M_i
  - $|M_{i\,R}|$, the number of named relations in M_i

3. EXPERIMENTS AND RESULTS

This study addressed four different classification problems: predicting the correctness of any particular semantic relation (*generic classifiers*) and predicting the correctness of a given type of semantic relation: ⊆, ⊇ or named (*specialized classifiers*). Note that the ⊥ relation has been discarded from our experiments due to the lack of negative examples. To address each of these problems, three different classifiers: the *J48* Decision Tree, the *NaiveBayes classifier*, and the *LibSVM* classifier, all of them provided by Weka [5] were used. Each classifier was applied using the whole set of attributes (Section 2) or a filtered set of attributes (*af*) obtained using a combination of the cfSubsetEval and the BestFirst algorithms [5]. To train and test the classifiers, each dataset was divided in the following way: approximately 70% of the data was used for training and 30% of the data was used for testing. This division was done manually to avoid the appearance of mappings coming from the same semantic relation in the training and the test sets. Note that the SW mappings coming from the same semantic relation share in common at least the first eight attributes, therefore, it is important to maintain them together in the same set (either the train or the test set) for a fair evaluation. To evaluate the classifiers and compare them against each other the following measures were selected: the percentage of correctly classified instances, the percentage of incorrectly classified instances and, the weighted average of the values obtained using the following measures for the positive and negative class: True Positives rate (TP), False Positives rate (FP), Precision, Recall, F-Measure (F-Mea) and ROC area value. More details about these measures can be found in [5]. The results obtained by the best classifier for each classification problem can be seen in Table 1.

Table 1. Best results obtained for each dataset

	Generic	**⊆**	**⊇**	**named**
	J48	**$J48_{af}$**	**NvBayes**	**J48**
Correct	74.2044%	85.2077%	98.0122%	76.1555%
Incorrect	25.7956%	14.7923%	1.9878%	23.8445%
TPRate	0.742	0.852	0.98	0.762
FPRate	0.254	0.122	0.06	0.209
Precision	0.76	0.889	0.984	0.79
Recall	0.742	0.852	0.98	0.762
F-Mea	0.747	0.851	0.981	0.766
ROC	0.749	0.875	0.995	0.767

4. CONCLUSIONS AND FUTURE WORK

In this paper, we investigate the problem of predicting the correctness of semantic relations. Our hypothesis is that ML methods can be adapted to exploit the SW as a source of knowledge to perform this task. The result of our experiments are promising, reaching a maximum of 74.2% of correctly classified semantic relations for classifiers able to validate the correctness of multiple types of semantic relations (*generic classifiers*) and up to 98% for classifiers focused on evaluating the correctness of one particular semantic relation (*specialized classifiers*).

Despite the success in the prediction process obtained by the classifiers, it is important to highlight that only 60% of the relations contained in these datasets were covered by the SW. This limits our approach to domains where semantic information is available, which constitutes an open problem for future research work.

5. REFERENCES

[1] Budanitsky, A. and Hirst, G. Evaluating WordNet-based measures of semantic distance. 2006. Computational Linguistics, 32(1):13-47.

[2] Calibrasi, R.L. and Vitanyi, P.M. The Google Similarity Distance. 2007. IEEE Transactions on Knowledge and Data Engineering, 19(3):370-383.

[3] Cimiano, P. 2006. Ontology Learning and Population from Text: Algorithms, Evaluation and Applications. Springer-Verlag New York, Inc.

[4] Sabou, M. and Gracia, J. Spider: Bringing Non-Equivalence Mappings to OAEI. 2008. In Proc. of the Third International Workshop on Ontology Matching

[5] Witten, I.H. and Eibe F. Data Mining. Practical Machine Learning Tools and Techniques. 2000. The Morgan Kaufmann Series in Data Management Systems