

Spatial Relation Extraction using Relational Learning

Parisa Kordjamshidi¹, Paolo Frasconi², Martijn Van Otterlo¹, Marie-Francine Moens¹, and Luc De Raedt¹

¹ Department of Computer Science, Katholieke Universiteit Leuven, Belgium

² Universit di Firenze, Italy

Abstract. The automatic extraction of spatial information is a challenging and novel task with many applications. We motivate our definition of this task and formulate it as an information extraction step prior to mapping to spatial semantics. Each sentence gives rise to several spatial relations between words representing *landmarks*, *trajectors* and *spatial indicators*. Learning to extract such spatial relations can be formulated as a typical relational classification problem, for which we employ the recently introduced kLog framework. We discuss modeling and representation, and show experimental results.

1 Background

An essential function of language is to express *spatial relationships* between objects and their relative location in the space. Understanding linguistic spatial descriptions is a challenging problem in robotics, navigation, human-machine interaction, query answering systems, etc [11]. The *automated* extraction of such information from natural language is very useful in many domains, and here we introduce a relational learning solution.

We identify two main abstract layers for extraction of spatial information [1, 8]: 1) a **linguistic** layer, which starts with unrestricted natural language and predicts the existence of spatial information at the sentence level, and identifies the words that play a particular spatial role; 2) a **formal** layer, in which the spatial roles are mapped onto spatial formal models [4]. For example, in “Give me the book on AI on the big table behind the wall.” a first step is to identify that a spatial relation (on) holds between book and table. This could then be mapped to a specific, *formal* relation (in some formalism) `AboveExternallyConnected(book, table)`.

Without considering formal relations here (for this, see [8]), we focus on the first (linguistic) level which is however motivated as a necessary prior step for mapping to formal spatial semantics. We call this task **spatial role labeling** (SpRL): *i) the identification of the words that play a role in describing spatial concepts, and ii) the classification of the role that they carry in the context of a spatial configuration.* The spatial roles are:

Trajector. The entity whose location or position is described. It can be static or dynamic; persons, objects, or events (also: *local/figure object, locatum*).

Landmark. The reference entity in relation to which the location or the motion of the trajector is specified. (also: *reference object* or *relatum*).

Spatial indicator. The element that defines constraints on spatial properties like the location of the trajector with respect to the landmark. It determines the type of spatial relation and is often expressed by a preposition, but can also be a verb, noun, adjective, or adverb.

The words can take part in one or more *spatial relations* that are expressed by the sentence. In the above mentioned sentence the location of *book* is described referring to the *table*, therefore *book* is a trajector and *table* is a landmark. Yet, *table* also has a relation with *wall*, but now with a different role. The words *on* and *behind* are spatial indicators and express the linguistic type of spatial relation. Now, $\{ on(book, table), behind(table, wall) \}$ is the set of spatial relations expressed by the sentence, and that $behind(book, wall)$ could be derived (cf. [12, 9]). Note that $on(book, AI)$ is not a spatial relation since this sense of *on* is not spatial. SpRL has not been studied systematically before. Often a restricted language is used to extract application-dependent relations and usually one focuses on phrases of which it is known that spatial information is present [7, 6, 11, 10]. And, even though semantic roles and the structure of the sentence given by (dependency) parsers contain useful information about spatial roles, these alone are not enough to directly map to spatial roles and relations. In the rest of this paper we describe a relational learning formulation of SpRL in the novel kLog framework, which is based on graph kernels and allows describing the learning problem in a declarative way. We discuss the imbalanced class problem (wrt. too many negative examples) and show results in 2 different settings.

2 Relational Learning for SpRL

The *input* of SpRL is natural language sentence S , which is a sequence of N words, $S = \langle w_1, w_2, \dots, w_N \rangle$. The words in the sentence have a number of local properties and also relationships to each other. The *output* is a set of spatial relations. We define a spatial relation as a triple $sr(SI, TR, LM)$, where the spatial indicator $SI = w_i$, the trajector $TR = w_j$, the landmark $LM = w_k$, $i, j, k \in [1, N]$ and $i \neq j \neq k$. We assume the roles are assigned to words. We now describe the kLog domain representation and learning model.

2.1 Exploiting relational structure using kLog

kLog [3]³ is a language that allows users to specify a relational database, domain knowledge and a target relational learning problem in a declarative way. The data model is based on representing *entities* and *relationships*. kLog *signatures* present the format of each table in the relational database. In this way the structure of the data is naturally presented as an entity-relationship (E/R) diagram and

³ <http://www.dsi.unifi.it/~paolo/ps/klog.pdf>

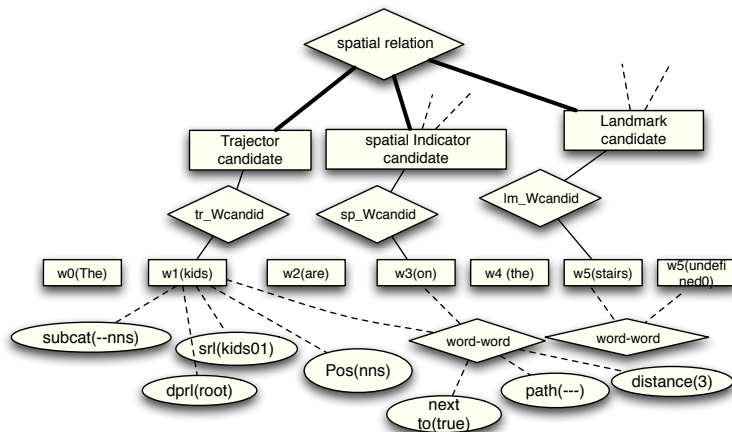


Fig. 1. Applied relational input features

features can be derived from that diagram. The learning setting then is *learning from interpretations* and each interpretation is a set of ground atoms.

Input: local features. The main entity of our model is a word and to obtain the properties of each word we perform a preprocessing step using the Charniak parser⁴ [2] and the LTH⁵ tool, producing semantic roles and other features in the CoNLL-2008 output format⁶. The local features include the word itself, part-of-speech tag, the dependency relation of the word to its syntactic head in the sentence, the semantic role and the subcategorization of the word. In kLog, an identifier is assigned to each word and the words and their properties are stored in a relational database. The word property tables are presented with their signatures such as in:

signature $word(w_id::self, word::property)::extensional$. e.g. $word(w0, the)$. $word(w1, kids)$.
signature $pos(w_id::word, posTag::property)::extensional$. e.g. $pos(w0, dt)$. $pos(w1, nns)$.

...

Signature is a reserved term used for defining each table, w_id is the identifier of each word, and *extensional* means the content of these tables are provided to kLog, while *intentional* would mean that the tables should be computed according to a given background knowledge.

Input: relational features. Feature analysis indicates that some pairwise relational features positively influence the learning model. These include the *path* in the parse tree between words, the binary *linear position* of the words wrt. each other, e.g. *before* or not, and the *distance*, i.e. the ratio of nodes on the path between two words wrt. the number of all the nodes in the parse tree.

⁴ <http://www.cog.brown.edu/~mj/Software.htm>

⁵ <http://barbar.cs.lth.se:8081/>

⁶ <http://barcelona.research.yahoo.net/dokuwiki/doku.php?id=conll2008:format>

Output: spatial relations. Each sentence is associated with a set of positive relations in the form of $sr(SI, TR, LM)$. The only issue is that trajectors and landmarks could be *implicit* and deleted by semantic ellipsis. For example, in “*There are red umbrellas on the right.*”, *on* (on the right) is spatial indicator *umbrellas* (red umbrellas) is trajector, and the landmark is undefined). Here we assume each spatial relation requires an explicit spatial indicator but the trajectors and landmarks can be undefined. To deal with implicit roles, we add a dummy word *undefined0* for these cases. kLog’s approach is to produce the candidate triples according to the background knowledge and learn to classify them using the relational(contextual) features of their components and predicts a target table of spatial relations:

signature mytarget(w1::sp_ind_can, w2::trajector_can, w3::landmark_can)::intensional.

This implies that we assign the roles of trajector, landmark and spatial indicator jointly and a sentence can naturally produce multiple spatial relations.

Background knowledge and candidate words. We employ background knowledge to guide the construction of possible candidate spatial relation triplets. Every preposition is assumed a candidate for being a spatial indicator, and nouns are considered as candidate trajectors and landmarks. The advantage of a declarative language such as kLog is that candidate words and their relations are easily produced by providing logical descriptions through the intentional tables. For example a trajector candidate is defined as:

signature trajector_can(tr_id::self)::intensional.

trajector_can(TID) :-word(W,-),(pos(W,'nn' ; pos(W,'nns');word(W,'undefined0')), atomic_concat(t,W,TID). Thus, a candidate trajector is either a noun or undefined0. To avoid too many features and too large related graphs, relational features are produced only for candidate words. Again, relational features can be directly programmed and extracted from the database.

Graphicalization. Graphicalization is the process of constructing a ground E/R diagram associated with a given interpretation. Fig 1 shows a part of a graph related to one sentence. There is one vertex for each tuple, labeled by the tuple itself and edges connect the entities and the relationships that have a common part in their identifiers. The graphs are turned into feature vectors using a graph kernel, which leads to a propositional learning problem at the end.

2.2 Dealing with negative examples

One of the main challenges of relational learning is the huge number of negative candidate relations compared to positives. To avoid building a biased model one way is to build a model based on a balanced data (i.e. ignoring the prior). Although the number of negatives can be reduced using background knowledge yet in many tasks getting a balanced data set is not possible. It will be very effective if the model is trained on the most confusing negatives, this will guarantee the robustness of the model in the future. However, we do not handle finding best negatives yet, but we choose them randomly from the possible candidates. We suggest a two phase evaluation to obtain a realistic evaluation of the trained model then. This will be discussed in the next section.

3 Experiments

Data set Our corpus ⁷consists of textual descriptions of 613 images taken from the IAPR TC-12 Image data set [5], denoted *CLEF*. It induces 1213 English sentences and 1716 corresponding spatial relations. We set a number of experiments here and show partial results, moreover we point to the challenge of negative examples for learning relations and the evaluation.

Experiment 1: Here we take a standard classification learning setting, producing candidate triples for both test and training sets according to background knowledge and we employ 10-fold cross-validation. The *precision* is 69.67%, *recall* 66.17% and *F1-measure* 67.87. We noticed two disadvantages that lead to learning a biased model and underestimating the performance that can be obtained by learning from the available data. Firstly, even though we use background knowledge the data is highly imbalanced. Second, a number of positives in the training phase is ignored because they were not picked as candidates by the background knowledge. Combined with a general lack of positives (wrt. negatives) this hurts performance. Ideally we would like to use all positives; hence our second experiment.

Experiment 2: Here we use *all* positives and select a number of negatives according to the background knowledge. We train a model based on the (balanced) data and evaluate with 10-fold cross validation. The *precision* is 98.46%, *recall* 93.41% and *F1-measure* 95.87%. Note that this is an overestimation, because the model is tested unrealistically regarding to the negatives that it will receive in reality when faced with new sentences. Hence to make sure that the trained model on the balanced data is robust enough a second phase of evaluation is performed by testing the model on all the negatives that have not been selected for training the model. The precision/recall can be computed based on summing up the contingency table of cross validation and the one resulted from a test on negatives, which gives *precision* 88%, *recall* 93.41% and *F1-measure* 90%. This shows a highly accurate model and the numbers are reliable for the performance of the system in the future.

Experiment 3. Another interesting experiment is to classify individual roles first and then produce the relations using a heuristic to show how it compares to joint learning of the relations. The F1-measure of classification of trajectors, landmarks and spatial indicators using kLog is (0.83, 0.79, 0.91) respectively, which outperforms previous results using a maximum entropy model. In those experiments we also employed a linear chain CRF model and results of classification of individual roles were not significantly different from kLog's. This indicates that the contextual and relational features in kLog give similar results as considering correlations between outputs in CRF sequence tagging if the influence of the difference between underlying models is assumed as trivial. However in the CRF context we employ a heuristic to produce relations from taggings, and this resulted in a less accurate extraction. A similar pipelining strategy in kLog itself performs equally poor. This is an expected outcome since there could be

⁷ The data sets will be made publicly available.

multiple relations per sentence and words can have multiple labels with respect to different indicators. Joint classification of the whole relation, which is easily modeled in kLog, is more effective.

4 Conclusion

The results for learning to extract spatial roles from natural language sentences are presented. We have employed the relational learning framework of kLog to classify candidate triplets of words representing spatial relations. To deal with the unbalanced data problem, learning from balanced data is performed and we suggested a two phase evaluation of the model. The performance of learning the roles jointly is higher compared to our previous problem formulation using sequence tagging. In future work we will map the spatial relations to more formal spatial semantics and spatial ontologies using kLog.

References

1. J. A. Bateman. Language and space: a two-level semantic approach based on principles of ontological engineering. *International Journal of Speech Technology*, 13(1):29–48, 2010.
2. E. Charniak and M. Johnson. Coarse-to-fine n-best parsing and maxent discriminative reranking. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 173–180, 2005.
3. P. Frasconi, Costa F., De Raedt L., and De Grave K. klog - a language for logical and relational learning with kernels. In *Technical report*, 2011.
4. Antony Galton. Spatial and temporal knowledge representation. *Journal of Earth Science Informatics*, 2(3):169–187, 2009.
5. Michael Grubinger, Paul Clough, Henning Müller, and Thomas Deselaers. The IAPR benchmark: A new evaluation resource for visual information systems. In *International Conference on Language Resources and Evaluation (LREC)*, 2006.
6. J. D. Kelleher. *A Perceptually Based Computational Framework for the Interpretation of Spatial Language*. PhD thesis, School of Computing Dublin City University, 2003.
7. T. Kollar, S. Tellex, D. Roy, and N. Roy. Toward understanding natural language directions. In *Proceeding of the 5th ACM/IEEE International Conference on Human-robot Interaction*, HRI '10, pages 259–266. ACM, 2010.
8. P. Kordjamshidi, M van Otterlo, and M. F. Moens. From language towards formal spatial calculi. In *Workshop on Computational Models of Spatial Language Interpretation (CoSLI 2010, at Spatial Cognition 2010)*, 2010.
9. P. Kordjamshidi, M van Otterlo, and M. F. Moens. Spatial role labeling: Task definition and annotation scheme. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, 2010.
10. H. Li, T. Zhao, S. Li, and J. Zhao. The extraction of trajectories from real texts based on linear classification. In *Proceedings of NODALIDA 2007 Conference*, pages 121–127, 2007.
11. D. A. Tappan. *Knowledge-Based Spatial Reasoning for Automated Scene Generation from Text Descriptions*. PhD thesis, 2004.
12. J. Zlatev. Spatial semantics. In *Hubert Cuyckens and Dirk Geeraerts (eds.) The Oxford Handbook of Cognitive Linguistics, Chapter 13*, pages 318–350, 2007.