SemEval-2012 Task 3: Spatial Role Labeling

Parisa Kordjamshidi
Katholieke Universiteit Leuven
parisa.kordjamshidi@cs.kuleuven.be

Steven Bethard
University of Colorado
steven.bethard@colorado.edu

Marie-Francine Moens
Katholieke Universiteit Leuven
sien.moens@cs.kuleuven.be

Abstract

This SemEval2012 shared task is based on a recently introduced spatial annotation scheme called Spatial Role Labeling. The Spatial Role Labeling task concerns the extraction of main components of the spatial semantics from natural language: trajectors, landmarks and spatial indicators. In addition to these major components, the links between them and the general-type of spatial relationships including region, direction and distance are targeted. The annotated dataset contains about 1213 sentences which describe 612 images of the CLEF IAPR TC-12 Image Benchmark. We have one participant system with two runs. The participant's runs are compared to the system in (Kordjamshidi et al., 2011c) which is provided by task organizers.

1 Introduction

One of the essential functions of natural language is to talk about spatial relationships between objects. The sentence *"Give me the book on AI on the big table behind the wall."* expresses information about the *spatial configuration* of the objects (book, table, wall) in some *space*. Particularly, it explains the region occupied by the *book* with respect to the *table* and the direction (orientation) of the *table* with respect to the *wall*. Understanding such spatial utterances is a problem in many areas, including robotics, navigation, traffic management, and query answering systems (Tappan, 2004).

Linguistic constructs can express highly complex, relational structures of objects, spatial relations between them, and patterns of motion through space relative to some reference point. Compared to natural language, formal spatial models focus on one particular spatial aspect such as orientation, topology or distance and specify its underlying spatial logic in detail (Hois and Kutz, 2008). These formal models enable spatial reasoning that is difficult to perform on natural language expressions.

Learning how to map natural language spatial information onto a formal representation is a challenging problem. The complexity of spatial semantics from the cognitive-linguistic point of view on the one hand, the diversity of formal spatial representation models in different applications on the other hand and the gap between the specification level of the two sides has led to the present situation that no well-defined framework for automatic spatial information extraction exists that can handle all of these aspects.

In a previous paper (Kordjamshidi et al., 2010b), we introduced the task of spatial role labeling (SpRL) and proposed an annotation scheme that is language-independent and practically facilitates the application of machine learning techniques. Our framework consists of a set of spatial roles based on the theory of holistic spatial semantics (Zlatevl, 2007) with the intent of covering the main aspects of spatial concepts at a course level, including both static and dynamic spatial semantics. This shared task is defined on the basis of that annotation scheme. Since this is the first shared task on the spatial information and this particular data, we proposed a simplified version of the original scheme. The intention of this simplification was to make this practice feasible in the given timeframe. However,

*First Joint Conference on Lexical and Computational Semantics (*SEM)*, pages 365–373,
Montréal, Canada, June 7-8, 2012.

the current task is very challenging particularly for learning the spatial links and relations.

The core problem of SpRL is: i) the *identification* of the words that play a role in describing spatial concepts, and ii) the *classification* of the relational *role* that these words play in the spatial configuration.

For example, consider again the sentence *"Give me the book on AI on the big table behind the wall."*. The phrase headed by the token *book* is referring to a trajector object. The trajector (TR) is an entity whose location is described in the sentence. The phrase headed by the token *table* is referring to the role of a landmark (LM). The landmark is a reference object for describing the location of a trajector. These two spatial entities are related by the spatial expression *on* denoted as spatial indicator (SP). The spatial indicator (often a preposition in English, but sometimes a verb, noun, adjective, or adverb) indicates the existence of spatial information in the sentence and establishes the type of a spatial relation. The spatial relations that can be extracted from the whole sentence are $<on_{SP}\ book_{TR}\ table_{LM}>$ and $<behind_{SP}\ table_{TR}\ wall_{LM}>$. One could also use spatial reasoning to *infer* that the statement <*behind book wall*> holds, however, such inferred relations are not considered in this task. Although the spatial indicators are mostly prepositions, the reverse may not hold- for example, the first preposition *on* only states the topic of the book, so <*on book AI*> is not a spatial relation. For each of the true spatial relations, a general type is assigned. The $<on_{SP}\ book_{TR}\ table_{LM}>$ relation expresses a kind of topological relationship between the two objects and we assign it a general type named *region*. The $<behind_{SP}\ table_{TR}\ wall_{LM}>$ relation expresses directional information and we assign it a general type named *direction*.

In general we assume two main abstraction layers for the extraction of spatial information (Bateman, 2010; Kordjamshidi et al., 2010a; Kordjamshidi et al., 2011a): (a) a **linguistic** layer, corresponding to the annotation scheme described above, which starts with unrestricted natural language and predicts the existence of spatial information at the sentence level by identifying the words that play a particular spatial role as well as their spatial relationship; (b) a **formal** layer, in which the spatial roles are mapped onto a spatial calculus model (Galton, 2009). For example, the linguistic layer recognizes that the spatial relation (*on*) holds between *book* and *table*, and the formal layer maps this to a specific, formal spatial representation, e.g., a logical representation like $\mathtt{AboveExternallyConnected(book, table)}$ or a formal qualitative spatial representation like `EC` (externally connected) in the RCC model (Regional Connection Calculus) (Cohn and Renz, 2008).

In this shared task we focus on the first (linguistic) level which is a necessary step for mapping natural language to any formal spatial calculus. The main roles that are considered here are trajector, landmark, spatial indicator, their links and the general type of their spatial relation. The general type of a relation can be *direction*, *region* or *distance*.

2 Motivation and related work

Spatial role labeling is a key task for applications that are required to answer questions or reason about spatial relationships between entities. Examples include systems that perform text-to-scene conversion, generation of textual descriptions from visual data, robot navigation tasks, giving directional instructions, and geographical information systems (GIS). Recent research trends (Ross et al., 2010; Hois et al., 2011; Tellex et al., 2011) indicate an increasing interest in the area of extracting spatial information from language and mapping it to a formal spatial representation. Although cognitive-linguistic studies have investigated this problem extensively, the computational aspect of making this bridge between language and formal spatial representation (Hois and Kutz, 2008) is still in its elementary stages. The possession of a practical and appropriate annotation scheme along with data is the first requirement. To obtain this one has to investigate and schematize both linguistic and spatial ontologies. This process needs to cover the necessary information and semantics on the one hand, and to maintain the practical feasibility of the automatic annotation of unobserved data on the other hand.

In recent research on spatial information and natural language, several annotation schemes have been proposed such as ACE, GUM, GML, KML, TRML which are briefly described and compared to SpatialML scheme in (MITRE Corporation, 2010). But to our knowledge, the main obstacles for employing machine learning in this context and the very limited usage of this effective approach have been (a) the lack of an agreement on a unique semantic model for spatial information; (b) the diversity of formal spatial relations; and consequently (c) the lack of annotated data on which machine learning can be employed to learn and extract the spatial relations. The most systematic work in this area includes the SpatialML (Mani et al., 2008) scheme which focuses on geographical information, and the work of (Pustejovsky and Moszkowicz, 2009) in which the pivot of the spatial information is the spatial verb. The most recent and active work is the ISO-Space scheme (Pustejovsky et al., 2011) which is based on the above two schemes. The ideas behind ISO-Space are closely related to our annotation scheme in (Kordjamshidi et al., 2010b), however it considers more detailed and fine-grained spatial and linguistic elements which makes the preparation of the data for machine learning more difficult.

Spatial information is directly related to the part of the language that can be visualized. Thus, the extraction of spatial information is useful for multimodal environments. One advantage of our proposed scheme is that it considers this dimension. Because it abstracts the spatial elements that could be aligned with the objects in images/videos and used for annotation of audio-visual descriptions (Butko et al., 2011). This is useful in the multimodal environments where, for example, natural language instructions are given to a robot for finding the way or objects.

Not much work exists on using annotations for learning models to extract spatial information. Our previous work (Kordjamshidi et al., 2011c) is a first step in this direction and provides a domain independent linguistic and spatial analysis to this problem. This shared task invites interested research groups for a similar effort. The idea behind this task is firstly to motivate the application of different machine learning approaches, secondly to investigate effective features for this task, and thirdly to reveal the practical problems in the annotation schemes and the annotated concepts. This will help to enrich the data and the annotation in parallel with the machine learning practice.

3 Annotation scheme

As mentioned in the introduction, the annotation of the data set is according to the general spatial role labeling scheme (Kordjamshidi et al., 2010b). The below example presents the annotated elements in this scheme.

> A woman$_{TR}$ and a child$_{TR}$ are walking$_{MOTION}$ over$_{SP}$ the square$_{LM}$.
>
> General-type: region
> Specific type: RCC
> Spatial value: PP (proper part)
> Dynamic
> Path: middle
> Frame of reference: –

According to this scheme the main spatial roles are,

Trajector (TR). The entity, i.e., person, object or event whose location is described, which can be static or dynamic; (also called: *local/figure object*, *locatum*). In the above example *woman* and *child* are two trajectors.

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

Spatial indicator (SP). The element that defines constraints on spatial properties such as the location of the trajector with respect to the landmark. The spatial indicator determines the type of spatial relation. The preposition *over* is annotated as the spatial indicator in the current example.

Moreover, the links between the three roles are annotated as a spatial **Relation**. Since each spatial relation is defined with three arguments we call it a **spatial triplet**. Each triplet indicates a **relation** between the three above mentioned spatial roles. The sentence contains two spatial relations of $<\mathit{over}_{SP}\ \mathit{woman}_{TR}\ \mathit{square}_{LM}>$ and $<\mathit{over}_{SP}\ \mathit{child}_{TR}\ \mathit{square}_{LM}>$, with the same spatial attributes listed below the example. In spatial information theory the relations and properties are usually grouped into the domains of *topological*, *directional*, and *distance* relations and also shape (Stock, 1997). Accordingly, we propose a mapping between the extracted spatial triplets to the coarse-grained type of spatial relationships including **region**, **direction** or **distance**. We call these types as **general-type** of the spatial relations and briefly describe these below:

Region. refers to a region of space which is always defined in relation to a landmark, e.g. the interior or exterior, e.g. *"the flower is in the vase"*.

Direction. denotes a direction along the axes provided by the different frames of reference, in case the trajector of motion is not characterized in terms of its relation to the region of a landmark, e.g. *"the vase is on the left"*.

Distance. states information about the spatial distance of the objects and could be a qualitative expression such as *close*, *far* or quantitative such as *12 km*, e.g. *"the kids are close to the blackboard"*.

The general-type of the relation in the example is annotated as *region*.

After extraction of these relations a next fine-grained step will be to map each general spatial relationship to an appropriate spatial calculi representation. This step is not intended for this task and the additional tags in the scheme will be considered in the future shared tasks. For example Region Connection Calculus RCC-8 (Cohn and Renz, 2008) representation reflects region-based topological relations. Topological or region-based spatial information has been researched in depth in the area of qualitative spatial representation and reasoning. We assume that the trajectors and landmarks can often be interpreted as spatial regions and, as a consequence, their relation can be annotated with a specific RCC-8 relation. The RCC type in the above example is specifically annotated as the PP (proper part). Similarly, the *direction* and *distance* relations are mapped to more specific formal representations.

Two additional annotations are about motion verbs and dynamism. Dynamic spatial information are associated with spatial movements and spatial changes. In dynamic spatial relations mostly motion verbs are involved. Motion verbs carry spatial information and influence the spatial semantics. In the above example the spatial indicator *over* is related to a motion verb *walking*. Hence the spatial relation is dynamic and *walking* is annotated as the *motion*. In contrast to the dynamic spatial relations, the static ones explain a static spatial configuration such as the example of the previous section $<on_{SP}$ $book_{TR}$ $table_{LM}>$.

In the case of dynamic spatial information a *path* is associated with the location of the trajector. In our scheme the *path* is characterized by the three values of *beginning*, *middle*, *end* and *zero*. The frame of reference can be intrinsic, relative or absolute and is typically relevant for directional relations. For more details about the scheme, see (Kordjamshidi et al., 2010b).

4 Tasks

The SemEval-2012 shared task is defined in three parts.

- The first part considers labeling the spatial indicators and trajector(s) / landmark(s). In other words at this step we consider the extraction of the individual roles that are tagged with TRAJECTOR, LANDMARK and SPATIAL_INDICATOR.

- The second part is a kind of relation prediction task and the goal is to extract triples containing (spatial-indicator, trajector, landmark). The prediction of the tag of RELATION with its three arguments of SP, TR, LM at the same time is considered.

- The third part concerns the classification of the type of the spatial relation. At the most coarse-grained level this includes labeling the spatial relations i.e. the triplets of (spatial indicator, trajector, landmark) with region, direction, and distance labels. This means the **general-type** of the RELATION should be predicted. The **general-type** is an attribute of the RELATION tag, see the example represented in XML format in section 5.1.

5 Preparation of the dataset

The annotated corpus that we used for this shared task is a subset of IAPR TC-12 image Benchmark (Grubinger et al., 2006). It contains 613 text files that include 1213 sentences in total. This is an extension of the dataset used in (Kordjamshidi et al., 2011c). The original corpus was available free of charge and without copyright restrictions. The corpus contains images taken by tourists with descriptions in different languages. The texts describe objects, and their absolute and relative positions in the image. This makes the corpus a rich resource for spatial information. However the descriptions are not always limited to spatial information. Therefore they are less domain-specific and contain free explanations about the images. Table 1 shows the detailed statistics of this data. The average length of the sentences in this data is about 15 words including punctuation marks with a standard deviation of 8.

The spatial roles are assigned both to phrases and their headwords, but only the **headwords** are evaluated for this task. The spatial relations indicate a triplet of these roles. The general-type is assigned to each triplet of spatial indicator, trajector and landmark.

At the starting point two annotators including one task-organizer and another non-expert annotator, annotated 325 sentences for the spatial roles and relations. The purpose was to realize the disagreement points and prepare a set of instructions in a way to achieve highest-possible agreement. From the first effort an inter-annotator agreement (Carletta, 1996) of 0.89 for Cohen's kappa was obtained. We continued with the a third annotator for the remaining 888 sentences. The annotator had an explanatory session and received a set of instructions and annotated examples to decrease the ambiguity in the annotations.

To avoid complexity only the relations that are directly expressed in the sentence are annotated and spatial reasoning was avoided during the annotations. Sometimes the trajectors and landmarks or both are implicit, meaning that there is no word in the sentence to represent them. For example in the sentence *Come over here*, the trajector *you* is only implicitly present. To be consistent with the number of arguments in spatial relations, in these cases we use the term *undefined* for the implicit roles. Therefore, the spatial relation in the above example is $<$*over*$_{SP}$ *undefined*$_{TR}$ *here*$_{LM}$$>$.

5.1 Data format

The data is released in XML format. The original textual files are split into sentences. Each sentence is placed in a <SENTENCE/> tag and assigned an identifier. This tag contains all the other tags which describe the content and spatial relations of one sentence.

The content of the sentence is placed in the <CONTENT/> tag. The words in each sentence are assigned identifiers depending on their specific roles. Trajectors, landmarks and spatial indicators are identified by <TRAJECTOR/>, <LANDMARK/> and <SPATIAL_INDICATOR/> tags, respectively. Each of these XML elements has an "ID" attribute that identifies a related word by its index. The "ID" prefixed by either "TW", "LW" or "SW", respectively for the mentioned roles. For example, a trajector with ID="TW2" corresponds to the word at index 2 in the sentence. Indexes start at 0. Commas, parentheses and apostrophes are also counted as tokens.

Spatial relations are assigned identifiers too, and relate the role-playing words to each other. Spatial relations are identified by the <RELATION/> tag. The spatial indicator, trajector and landmark for the relation are identified by the "SP", "TR" and "LM" attributes, respectively. The values of these attributes correspond to the "ID" attributes in the <TRAJECTOR/>, <LANDMARK/> and <SPATIAL_INDICATOR/> elements. If a trajector or landmark is implicit, then the index of "TR" or "LM" attribute will be set to a dummy index. This dummy index is equal to the index of the last word in the sentence plus one. In this case, the value of TRAJECTOR or LANDMARK is set to "undefined". The coarse-grained spatial type of the relation is indicated by the "GENERAL_TYPE" attribute and gets one value in {REGION, DIRECTION, DISTANCE}. In the original data set there are cases annotated with multiple spatial types. This is due to the ambiguity and/or under-specificity of natural language compared to formal spatial representations (Kordjamshidi et al., 2010a). In this task the general-type with a higher priority by the annotator is provided. Here, by the high priority type, we mean the general type which has been the most informative

	Spatial Roles			Relations	General Types		
Sentences	TR	LM	SP	Spatial triplets	Region	Direction	Distance
1213	1593	1408	1464	1715	1036	644	35

Table 1: Number of annotated components in the data set.

and relevant type for a relation, from the annotator's point of view. This task considers labeling words rather than phrases for all spatial roles. However, in the XML file for spatial indicators often the whole phrase is tagged. In these cases, the index of the indicator refers to one word which is typically the spatial preposition of the phrase. For evaluation only the indexed words are compared and should be predicted correctly.

Below is one example copied from the data. For more examples and details about the general annotation scheme see (Kordjamshidi et al., 2010b).

<SENTENCE ID='S11'>
<CONTENT >
there are red umbrellas in a park on the right .
</CONTENT>
<TRAJECTOR ID='TW3'>
umbrellas
</TRAJECTOR>
<LANDMARK ID='LW6'>
park
</LANDMARK>
<SPATIAL_INDICATOR ID='SW4'>
in
</SPATIAL_INDICATOR>
<RELATION ID='R0' SP='SW4' TR='TW3' LM='LW6' GENERAL_TYPE='REGION'/>
<SPATIAL_INDICATOR ID='SW7'>
on the right
</SPATIAL_INDICATOR>
<RELATION ID='R1' SP='SW7' TR='TW3' LM='LW6' GENERAL_TYPE='DIRECTION'/>
</SENTENCE>

The dataset, both train and test, also the 10-fold splits are made available in the LIIR research group webpage of KU Leuven.[1]

6 Evaluation methodology

According to the usual setting of the shared tasks our evaluation setting was based on splitting the data set into a training and a testing set. Each set contained about 50% of the whole data. The test set released without the ground-truth labels. However, after the systems submission deadline the ground-truth test was released. Hence the participant group performed an additional 10-fold cross validation evaluation too. We report the results of both evaluation settings.

Prediction of each component including TRAJECTORS, LANDMARKS and SPATIAL-INDICATORS is evaluated on the test set using their individual spatial element XML tags. The evaluation metrics of precision, recall and F1-measure are used, which are defined as:

$$recall = \frac{TP}{TP+FN} \tag{1}$$

$$precision = \frac{TP}{TP+FP} \tag{2}$$

$$F1 = \frac{2*recall*precision}{(recall+precision)}, \tag{3}$$

where:

TP = the number of system-produced XML tags that match an annotated XML tag,
FP = the number of system-produced XML tags that do not match an annotated tag,
FN = the number of annotated XML tags that do not match a system-produced tag.

For the roles evaluation two XML tags match when they have exactly same identifier. In fact, when the identifiers are the same then the role and the word index are the same. In addition, systems are evaluated on how well they are able to retrieve triplets of (trajector, spatial-indicator, landmark), in terms of precision, recall and F1-measure. The TP, FP, FN are counted in a similar way but two RELATION tags match if the combination of their TR, LM and SP is exactly the same. In other words a true prediction requires all the three elements are correctly predicted at the same time.

The last evaluation is on how well the systems are able to retrieve the relations and their general type

[1]http://www.cs.kuleuven.be/groups/liir/software/SpRL_Data/

i.e {region, direction, distance} at the same time. To evaluate the GENERAL-TYPE similarly the RELATION tag is checked. For a true prediction, an exact match between the ground-truth and all the elements of the predicted RELATION tag including TR, LM,SP and GENERAL-TYPE is required.

7 Systems and results

One system with two runs was submitted from the University of Texas Dallas. The two runs (Roberts and Harabagiu, 2012), UTDSPRL-SUPERVISED1 and UTDSPRL-SUPERVISED2 are based on the joint classification of the spatial triplets in a binary classification setting. To produce the candidate (indicator, trajector, landmark) triples, in the first stage heuristic rules targeting a high recall are used. Then a binary support vector machine classifier is employed to predict whether a triple is a spatial relation or not. Both runs start with a large number of manually engineered features, and use floating forward feature selection to select the most important ones. The difference between the two runs of UTDSPRL-SUPERVISED1 and UTDSPRL-SUPERVISED2 is their feature set. Particularly, in UTDSPRL-SUPERVISED1 a joint feature based on the conjunctions (e.g. *and*, *but*) is considered before running feature selection but this feature is removed in UTDSPRL-SUPERVISED2.

The submitted runs are compared to a previous system from the task organizers (Kordjamshidi et al., 2011c) which is evaluated on the current data with the same settings. This system, KUL-SKIP-CHAIN-CRF, uses a skip chain conditional random field (CRF) model (Sutton and MacCallum, 2006) to annotate the sentence as a sequence. It considers the long distance dependencies between the prepositions and nouns in the sentence.

The type and structure of the features used in the UTD and KUL systems are different. In the UTD system, the classifier works on triples and the features are of two main types: (a) argument-specific features about the trajector, landmark, or indicator e.g., the landmark's hypernyms, or the indicator's first token; and (b) joint features that consider two or more of the arguments, e.g. the dependency path between indicator and landmark. For more detail, see (Roberts and Harabagiu, 2012). In the KUL system, the classifier works on all tokens in a sentence, and a number of linguistically motivated local and pairwise features over candidate words and prepositions are used. To consider long distance dependencies a template, called a preposition template, is used in the general CRF framework. Loopy belief propagation is used for inference. Mallet[2] and GRMM:[3] implementations are employed there.

Label	Precsion	Recall	F1
TRAJECTOR	0.731	0.621	0.672
LANDMARK	0.871	0.645	0.741
SPATIAL-INDICATOR	0.928	0.712	0.806
RELATION	0.567	0.500	0.531
GENERAL-TYPE	0.561	0.494	0.526

Table 2: UTDSPRL-SUPERVISED1: The University of Texas-Dallas system with a larger number of features,test/train one split.

Label	Precsion	Recall	F1
TRAJECTOR	0.782	0.646	0.707
LANDMARK	0.894	0.680	0.772
SPATIAL-INDICATOR	0.940	0.732	0.823
RELATION	0.610	0.540	0.573
GENERAL-TYPE	0.603	0.534	0.566

Table 3: UTDSPRL-SUPERVISED2: The University of Texas-Dallas system with a smaller number of features, test/train one split.

Label	Precsion	Recall	F1
TRAJECTOR	0.697	0.603	0.646
LANDMARK	0.773	0.740	0.756
SPATIAL-INDICATOR	0.913	0.887	0.900
RELATION	0.487	0.512	0.500

Table 4: KUL-SKIP-CHAIN-CRF: The organizers' system (Kordjamshidi et al., 2011c)- test/train one split.

Tables 2, 3 and 4 show the results of the three runs in the standard setting of the shared task using the train/test split. In this evaluation setting the UTDSPRL-SUPERVISED2 run achieves the highest performance on the test set, with F1 of 0.573 for the full triplet identification task, and an F1 of 0.566 for additionally classifying the triplet's general-type

[2]http://mallet.cs.umass.edu/download.php
[3]http://mallet.cs.umass.edu/grmm/index.php

System	Precsion	Recall	F1
KUL-SKIP-CHAIN-CRF	0.745	0.773	0.758
UTDSPRL-SUPERVISED2	0.773	0.679	0.723

Table 5: The RELATION extraction of KUL-SKIP-CHAIN-CRF (Kordjamshidi et al., 2011c) vs. UTDSPRL-SUPERVISED2 evaluated with 10-fold cross validation

correctly. It also consistently outperforms both the UTDSPRL-SUPERVISED1 run and the KUL-SKIP-CHAIN-CRF system on each of the individual trajector, landmark and spatial-indicator extraction.

The dataset was relatively small, so we released the test data and the two systems were additionally evaluated using 10-fold cross validation. The results of this cross-validation are shown in Table 5. The UTDSPRL-SUPERVISED2 run achieves a higher precision, while the KUL-SKIP-CHAIN-CRF system achieves a higher recall. It should be mentioned the 10-fold splits used by KUL and UTD are not the same. This implies that the results with exactly the same cross-folds may vary slightly from these reported in Table 5.

Using 10-fold cross validation, we also evaluated the classification of the general-type of a relation given the manually annotated positive triplets. The UTDSPRL-SUPERVISED2 system achieved F1= 0.974, and similar experiments using SMO-SVM in (Kordjamshidi et al., 2011b; Kordjamshidi et al., 2011a) achieved F1= 0.973. Thus it appears that identifying the general-type of a relation is a relatively easy task on this data.

Discussion. Since the feature sets of the two systems are different and given the evaluation results in the two evaluation settings, it is difficult to assert which model is better in general. Obviously using joint features potentially inputs richer information to the model. However, it can increase the sparsity in one hand and overfitting on the training data on the other hand. Another problem is that finding heuristics for high recall that are sufficiently general to be used in every domain is not an easy task. By increasing the number of candidates the dataset imbalance will increase dramatically. This can cause a lower performance of a joint model based on a binary classification setting when applied on different data sets. It seems that this task might require a more elaborated structured output prediction model which can consider the joint features and alleviate the problem of huge negatives in that framework while considering the correlations between the output components.

8 Conclusion

The SemEval-2012 spatial role labeling task is a starting point to formally consider the extraction of spatial semantics from the language. The aim is to consider this task as a standalone linguistic task which is important for many applications. Our first practice on this task and the current submitted system to SemEval 2012 clarify the type of the features and the machine learning approaches appropriate for it. The proposed features and models help to perform this task automatically in a reasonable accuracy. Although the spatial scheme is domain independent, the achieved accuracy is dependent on the domain of the used data for training a model. Our future plan is to extend the data for the next workshops and to cover more semantic aspects of spatial information particularly for mapping to formal spatial representation models and spatial calculus.

Acknowledgments

Special thanks to Martijn Van Otterlo for his great cooperation from the initiation phase and in the growth of this task. Many thanks to Sabine Drebusch for her kind and open cooperation in annotating the very first dataset. Thanks to Tigist Kassahun for her help in annotating the current dataset. Thanks the participant team of the University of Texas Dallas and their useful feedback on the annotated data.

References

J. A. Bateman. 2010. Language and space: a two-level semantic approach based on principles of ontological engineering. *International Journal of Speech Technology*, 13(1):29–48.

T. Butko, C. Nadeu, and A. Moreno. 2011. A multilingual corpus for rich audio-visual scenedescription in a meeting-room environment. In *ICMI workshop on multimodal corpora for machine learning: Taking Stock and Roadmapping the Future*.

J. Carletta. 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22(2):249–254.

A. G. Cohn and J. Renz. 2008. Qualitative spatial representation and reasoning. In *Handbook of Knowledge Representation*, volume 3 of *Foundations of Artificial Intelligence*, pages 551 – 596. Elsevier.

A. Galton. 2009. Spatial and temporal knowledge representation. *Journal of Earth Science Informatics*, 2(3):169–187.

M. Grubinger, P. Clough, Henning Müller, and Thomas Deselaers. 2006. The IAPR benchmark: A new evaluation resource for visual information systems. In *International Conference on Language Resources and Evaluation (LREC)*.

J. Hois and O. Kutz. 2008. Natural language meets spatial calculi. In Christian Freksa, Nora S. Newcombe, Peter Gärdenfors, and Stefan Wölfl, editors, *Spatial Cognition*, volume 5248 of *Lecture Notes in Computer Science*, pages 266–282. Springer.

J. Hois, R. J. Ross, J. D. Kelleher, and J. A. Bateman. 2011. Computational models of spatial language interpretation and generation. In *COSLI-2011*.

P. Kordjamshidi, M. van Otterlo, and M. F. Moens. 2010a. From language towards formal spatial calculi. In *Workshop on Computational Models of Spatial Language Interpretation (CoSLI 2010, at Spatial Cognition 2010)*.

P. Kordjamshidi, M. van Otterlo, and M. F. Moens. 2010b. Spatial role labeling: Task definition and annotation scheme. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*.

P. Kordjamshidi, J. Hois, M. van Otterlo, and M.-F. Moens. 2011a. Machine learning for interpretation of spatial natural language in terms of qsr. Poster Presentation at the 10th International Conference on Spatial Information Theory (COSIT'11).

P. Kordjamshidi, J. Hois, M. van Otterlo, and M.F. Moens. 2011b. Learning to interpret spatial natural language in terms of qualitative spatial relations. *Representing space in cognition: Interrelations of behavior, language, and formal models. Series Explorations in Language and Space, Oxford University Press, submitted*.

P. Kordjamshidi, M. Van Otterlo, and M.F. Moens. 2011c. Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Trans. Speech Lang. Process.*, 8:1–36, December.

I. Mani, J. Hitzeman, J. Richer, D. Harris, R. Quimby, and B. Wellner. 2008. SpatialML: Annotation scheme, corpora, and tools. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, and Daniel Tapias, editors, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA).

MITRE Corporation. 2010. SpatialML: Annotation scheme for marking spatial expression in natural language. Technical Report Version 3.0.1, The MITRE Corporation.

J. Pustejovsky and J.L. Moszkowicz. 2009. Integrating motion predicate classes with spatial and temporal annotations. In *CoLing 2008: Companion volume Posters and Demonstrations*, pages 95–98.

J. Pustejovsky, J. Moszkowicz, and M. Verhagen. 2011. Iso-space: The annotation of spatial information in language. In *Proceedings of ISA-6: ACL-ISO International Workshop on Semantic Annotation*.

K. Roberts and S.M. Harabagiu. 2012. Utd-sprl: A joint approach to spatial role labeling. In *Submitted to this workshop of SemEval-2012*.

R. Ross, J. Hois, and J. Kelleher. 2010. Computational models of spatial language interpretation. In *COSLI-2010*.

O. Stock, editor. 1997. *Spatial and Temporal Reasoning*. Kluwer.

C. Sutton and A. MacCallum. 2006. Introduction to conditional random fields for relational learning. In Lise Getoor and Ben Taskar, editors, *Introduction to Statistical Relational Learning*. MIT Press.

D. A. Tappan. 2004. *Knowledge-Based Spatial Reasoning for Automated Scene Generation from Text Descriptions*. Ph.D. thesis.

S. Tellex, T. Kollar, S. Dickerson, M. R. Walter, and N. Roy A. G. Banerjee, S. Teller. 2011. Understanding natural language commands for robotic navigation and mobile manipulation. In *Proceedings of the National Conference on Artificial Intelligence (AAAI), San Francisco, CA*.

J. Zlatevl. 2007. Spatial semantics. *In Hubert Cuyckens and Dirk Geeraerts (eds.) The Oxford Handbook of Cognitive Linguistics, Chapter 13*, pages 318–350.