Detecting Spatial and Motion Relations in Texts

Fadi Botros
University of Calgary
fnmbotro@ucalgary.ca

Eric Nichols
Honda Research Institute Japan Co. Ltd.
e.nichols@jp.honda-ri.com

1 Introduction

Understanding human language about location and motion is important for many applications including robotics, navigation systems, and wearable computing. To provide a framework for representing and detecting spatial and motion relations, Kordjamshidi et al. [6] proposed the task of Spatial Role Labeling (SpRL), and shared tasks have been organized for SemEval 2012, 2013, and 2015. In this paper we present `HRI-CRF-VW`, a system that conducts spatial role labeling in two phases: (1) it detects spatial relation argument and trigger expression candidates with a CRF sequential labeling model that uses a combination of distributed word representations and lexico-syntactic features; (2) given relation candidate tuples, it jointly classifies relations into types and labels the spatial roles of arguments with a multi-class classification model that uses a combination of syntactic and semantic features. Evaluation on SemEval 2013 test data shows that our system outperforms all known existing systems for the majority of spatial roles and outperforms all but one system on relation classification. Preliminary evaluation on SemEval 2015 data shows comparable performance despite a more challenging task setting.

2 Spatial and Motion Relation Detection

2.1 Spatial Role Labeling

Kordjamshidi et al. [7] proposed the task of Spatial Role Labeling (SpRL) to detect spatial and motion relations in text. SpRL was modeled after semantic role labeling, with spatial indicators taking the place of predicate to signal the presence of a relation, and the spatial roles in place of semantic roles.

A canonical example of a spatial relation from [7] is: *Give me the [grey book]*$_{TR}$ *[on]*$_{SP}$ *the [large table]*$_{LM}$. The `SPATIAL_INDICATOR` *on* indicates that there is a *spatial configuration* relation between the `TRAJECTOR` (primary object of spatial focus) and the `LANDMARK` (secondary object of spatial focus).

SpRL was formalized as a task of classifying $< w_{SP}, w_{TR}, w_{LM} >$ tuples as spatial relations or not.

2.2 SemEval 2012

The first shared task forSpRL was organized at SemEval 2012 [6] and included the following tasks:

- **Task I**: *simple spatial role* identification of `SPATIAL_INDICATOR`, `TRAJECTOR`, and `LANDMARK`
- **Task II**: binary classification of tuples into `RELATION` or `NO_REL`

The dataset used was the CLEF IAPR TC-12 Image Benchmark suite. It consists of 1,213 sentences describing 612 images annotated with *simple* spatial roles and relations. The data is described in detail in [6].

2.3 SemEval 2013

The SemEval 2013 SpRL shared task [5] continued in SemEval-2102's direction while adding a new dataset annotated with an extended set of spatial roles including motion information. The tasks included:

- **Task A**: *simple spatial role* identification identical to SemEval 2012 Task I
- **Task B**: spatial relation classification identical to SemEval 2012 Task II
- **Task C**: *extended spatial role* identification of `SPATIAL_INDICATOR`, `TRAJECTOR`, `LANDMARK`, `MOTION_INDICATOR`, `PATH`, `DIRECTION`, and `DISTANCE`

The dataset used in Tasks A and B was identical to SemEval 2012. For Task C, the Confluence Project Corpus was used. The data consists of descriptions of locations where latitude and longitude lines intersect. It contains a total of 2,105 annotations of extended spatial roles and relations across 1,422 sentences.

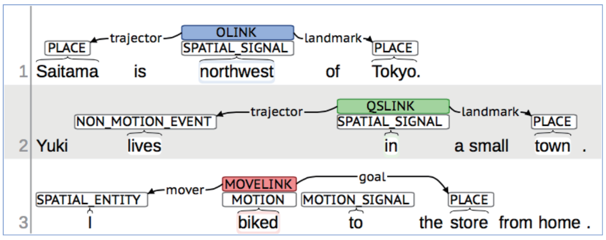

Figure 1: Example relations from SemEval 2015 task

2.4 SemEval 2015

The SpRL task was reformulated and reintroduced in SemEval 2015[1]. The biggest change was the decoupling of the semantic *category* and *role* of spatial relation arguments. A taxonomy of Spatial Element and Signal types was introduced to describe the semantic content of arguments independent of participation in relations, and spatial roles were treated as instance-specific annotations on spatial and motion relations.

The Spatial Element and Signal types introduced are: `SPATIAL_ENTITY`, `PATH`, `PLACE`, `MOTION`, `NON_MOTION_EVENT`, `MEASURE`, `SPATIAL_SIGNAL`, and `MOTION_SIGNAL`.

Spatial and motion relations were refactored into:

- `QSLINK`: qualitative spatial relation
- `OLINK`: spatial orientation relation
- `MOVELINK`: motion relation

The spatial roles that were introduced as annotations on relations include:

[1] http://alt.qcri.org/semeval2015/task8/

EF.1 Raw string in a 5-word window
(i.e. *Saitama is northwest of Tokyo*)
EF.2 Lemma in a 5-word window
(i.e. *Saitama be northwest of Tokyo*)
EF.3 POS in a 5-word window
(i.e. `NNP VBZ RB IN NNP`)
EF.4 Named Entity in a 5-word window
(i.e. `LOC NONE NONE NONE LOC`)
EF.5 Lemma concatenated with the POS in a 3-word window
(i.e *be*::`VBZ` *northwest*::`RB` *of*::`IN`)
EF.6 Named Entity concatenated with the POS in a 3-word window
(i.e `NONE::VBZ NONE::RB NONE ::IN`)
EF.7 Direct dependency on the head of the sentence if present
(i.e. *advmod*)
EF.8 Direct dependency on the head of the sentence concatenated with the lemma of the head
(i.e. *advmod*::*be*)
EF.9 300-dimension GloVe word vector
EF.10 POS bigrams for a 5-word window
(i.e. `NNP_VBZ VBZ_RB RB_IN IN_NNP`)
EF.11 Raw string n-grams for 3-word window
(i.e. *is_northwest northwest_of*)

Figure 2: Features for spatial element/signal detection

- `QSLINK` & `OLINK`: `TRIGGER`, `TRAJECTOR` and `LANDMARK`
- `MOVELINK`: `TRIGGER`, `MOVER`, and `GOAL`

The dataset for SemEval 2015 consists of portions of the corpora from past SemEval tasks as well as a new dataset consisting of passages from guidebooks. Following the schema described in this section, a total of 6,782 spatial elements and signals comprising 2,186 relations were annotated.

3 Related Research

3.1 Past SemEval Systems

3.1.1 KUL-SKIP-CHAIN-CRF

`KUL-SKIP-CHAIN-CRF` [7] was a skip-chain CRF-based sequential labeling model. It used a combination of lexico-syntactic information and semantic role information and employed a system called *preposition templates* to represent long distance dependencies. It was used as a baseline system in SemEval 2012 and 2013.

3.1.2 UTD-SpRL

`UTD-SpRL` [11] was an entry into the SemEval 2012 Spatial Role labeling task. The task setting was to identify spatial relations in texts, classify the relation type as either `REGION`, `DIRECTION` or `DISTANCE`, and label the role of each argument as `TRAJECTOR`, `LANDMARK`, or `INDICATOR`. `UTD-SpRL` adopted a joint relation detection and role labeling approach with the motivation that roles in spatial relations were dependent on each other. The approach used heuristics to gather spatial relation candidate tuples. A hand-crafted dictionary was used to detect `INDICATOR` candidates, and noun phrase heads were treated as `TRAJECTOR` and `LANDMARK` candidates. A model for relation classification and role labeling was then trained with libLINEAR using POS, lemma, and dependency-path-based features, with feature selection used to prune away ineffective features.

3.1.3 UNITOR-HMM-TK

`UNITOR-HMM-TK` [2] was an entry into the SemEval 2013 SpRL task. Its approach was to divide SpRL into two sub-tasks: (1) spatial annotation classification and (2) spatial relation identification.

`UNITOR-HMM-TK` adapted a sequential labeling approach to spatial annotation classification using SVM^{hmm}. Because spatial indicators were considered a closed class of expressions whose existence is a good indicator of presence of semantic relations, a pipeline approach was adopted with indicator detection followed by spatial role classification. In addition to indicator features, shallow grammatical features in the form of POS n-grams were used in place of richer syntactic information in order to avoid overfitting. The model also incorporated word space representations that were learned using singular value decomposition on matrices of PMI scores derived from cooccurrence counts.

Element Type	P	R	F1
Place	0.802	0.777	0.789
Spatial Entity	0.793	0.653	0.716
Spatial Signal	0.750	0.603	0.668
Motion	0.823	0.700	0.756
Motion Signal	0.766	0.600	0.673
Path	0.815	0.614	0.701
Non Motion Event	0.663	0.371	0.476
Measure	0.889	0.707	0.788
OVERALL	0.795	0.674	0.730

Table 1: `HRI-CRF-VW`'s spatial element/signal detection results, tested on the SemEval 2015 dataset

`UNITOR-HMM-TK`'s approach to spatial relation identification was to avoid feature engineering by employing an SVM model with a smoothed partial tree kernel over modified dependency trees to capture rich syntactic information.

3.2 Semantic Role Labeling

SpRL's task formulation was inspired by semantic role labeling – in particular the role labels of FrameNet [3] that are shared across predicates. It is thus unsurprising that SpRL approaches often takes inspiration from SRL. For an overview, see [8]. A state-of-the-art SRL system using phrase vectors is described in [4].

4 Spatial Element and Signal Detection

4.1 Approach

`HRI-CRF-VW` uses a feature-rich CRF labeling model to jointly label spatial elements, spatial and motion signals. Previous systems [7, 2] proposed a two-step sequential labeling method for this task. In the first step, they label spatial and motion signals since they indicate that there is a relation in the sentence. In the second step, they label all the other spatial arguments in the sentence using the extracted spatial and motions signals as features. However, any errors made in the first step will deteriorate the performance of the second. By combining the two steps, our system avoids this problem.

The CRF model labels each word in a sentence with a SemEval 2015 spatial element/signal, or with `NONE`. In line with `UNITOR-HMM-TK` [2], shallow lexico-syntactic features are applied instead of the full syntax of the sentence to avoid over-fitting over the training data. Word vectors are also used to capture the fine-grained lexical meaning.

For the detection of spatial elements, spatial signals, and motion signals, each word in Figure 1 is represented by the features in Figure 2.

4.2 Evaluation

4.2.1 Setup

Sentences were processed with the Stanford CoreNLP software[2] for POS tagging, lemmatization, NER, and dependency parsing. The word representations are 300-dimension GloVe [10] publicly-available[3] word vectors trained on 42 billion tokens of Web data. The model

[2] http://nlp.stanford.edu/software/corenlp.shtml
[3] http://www-nlp.stanford.edu/projects/glove/

	KUL-SKIP-CHAIN-CRF			UTDSpRL-SUPERVISED2			UNITOR-HMM-TK			HRI-CRF-VW		
Spatial Role	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
Spatial Indicator	0.913	0.887	0.900	0.940	0.732	0.823	**0.967**	0.889	0.926	0.965	**0.901**	**0.932**
Trajector	0.697	0.603	0.646	**0.782**	0.646	**0.707**	0.684	0.681	0.682	0.681	**0.689**	0.685
Landmark	0.773	0.740	0.756	**0.894**	0.680	0.772	0.741	**0.835**	0.785	0.869	0.789	**0.827**

Table 2: A comparison of HRI-CRF-VW to other systems from previous SemEval SpRL tasks. KUL-SKIP-CHAIN-CRF and UTDSpRL-SUPERVISED2 results are from SemEval 2012. UNITOR-HMM-TK results are from SemEval 2013. All systems were tested on the CLEF IAPR TC-12 dataset.

	UNITOR-HMM-TK			HRI-CRF-VW		
Spatial Role	**P**	**R**	**F1**	**P**	**R**	**F1**
Spatial Indicator	0.609	0.470	0.536	0.680	0.549	**0.608**
Motion Indicator	0.892	0.294	0.443	0.826	0.645	**0.724**
Trajector	0.565	0.317	0.406	0.687	0.533	**0.601**
Landmark	0.662	0.476	**0.554**	0.629	0.488	0.549
Path	0.775	0.295	0.427	0.676	0.600	**0.636**
Direction	0.312	0.229	0.264	0.701	0.445	**0.545**
Distance	0.946	0.331	0.490	0.824	0.635	**0.717**

Table 3: UNITOR-HMM-TK and HRI-CRF-VW results on SemEval 2013 - Task C

was trained using CRFsuite [9] with the L-BFGS optimization algorithm with L2 regularization and a delta value of 1e-5.

4.2.2 Datasets

Our system was trained and tested on the SemEval 2015 task data as described in Section 2.4.

4.3 Results

To evaluate our system, we tested it on the data provided for SemEval 2015, mentioned in Section 4.2.2. Table 1 outlines the results on SemEval 2015 data using 5-fold cross validation. Even though SpRL tasks from previous years had different annotation schemes, we also evaluated on data from previous SemEval tasks to compare our system with other systems. The first comparison test was done using SemEval 2012 - Task I's *simple* annotation scheme and dataset. Table 2 compares the results of all the systems on this task. The F1 scores show that HRI-CRF-VW outperforms all the other systems in SPATIAL_INDICATOR and LANDMARK classification while UTDSpRL-SUPERVISED2 leads in TRAJECTOR classification.

HRI-CRF-VW was further tested on SemEval 2013 - Task C since it had a more comparable annotation setting and dataset to SemEval 2015. Table 3 shows the results of the UNITOR-HMM-TK system and the HRI-CRF-VW system on this task. HRI-CRF-VW displays a significant increase in F1 score over the competing system in all spatial roles except for LANDMARK.

5 Spatial Relation Classification and Argument Labeling

5.1 Approach

To identify spatial relations, the HRI-CRF-VW system determines which spatial elements and signals, discovered in the previous classification step, can be combined to form valid spatial relations. Since the type of a relation (QSLINK, OLINK or MOVELINK) is dependent upon its arguments, our method, inspired by UTD-SpRL [11], jointly classifies spatial relations and labels participating arguments in one classification step.

First, triggers are extracted from each sentence. Triggers in every relation are either a SPATIAL_SIGNAL (for QSLINK and OLINK) or a MOTION (for MOVELINK). All possible candidate relations in a sentence are then generated using all the other spatial elements in the sentence. A candidate tuple consists of an extracted trigger and two other spatial elements in the sentence; arg1 and arg2. Each tuple is represented by three main groups of features outlined in Figure 3. We then apply a one-against-all multi-class classifier to classify each candidate relation tuple into one of three possible classes. Three separate classifiers are trained, one for each spatial relation type, using Vowpal Wabbit's [1] online stochastic gradient descent. The classes used by the QSLINK and OLINK classifiers are:

Features representing the extracted trigger:
RF.1 Raw string
RF.2 Lemma
RF.3 POS
RF.4 RF.2 concatenated with RF.3
Features representing each of the two arguments:
RF.5 Raw string
RF.6 Lemma
RF.7 POS
RF.8 RF.6 concatenated with RF.7
RF.9 Spatial element type (i.e Place, Path, etc.)
RF.10 RF.9 of each argument concatenated together
RF.11 RF.10 concatenated with RF.2
RF.12 Direction of the argument with the respect to the extracted trigger (i.e left/right)
RF.13 RF.12 of each argument concatenated together
RF.14 RF.13 concatenated with RF.2
RF.15 Boolean value representing whether there are other spatial elements in between the argument and the extracted trigger
RF.16 RF.15 of each argument concatenated together
RF.17 Dependency path between the argument and the extracted trigger (i.e. $\uparrow conj \downarrow dep \downarrow nsubj$)
RF.18 RF.17 of each argument concatenated together
RF.19 Dependency path between the two arguments
RF.20 Length of the dependency path between the argument and the extracted trigger
RF.21 Bag-of-words of tokens in between the argument and the extracted trigger
RF.22 Number of tokens in between the argument and the extracted trigger
RF.23 RF.22 of each argument added together
RF.24 Boolean value representing whether either of the arguments are null values
Features representing the spatial elements that are directly to the left and to the right of the trigger:
RF.25 Raw string
RF.26 Lemma
RF.27 POS
RF.28 RF.26 concatenated with RF.27
RF.29 Number of tokens in between the spatial element and the extracted trigger

Figure 3: Features for joint spatial relation classification and role labeling

Class 1 - arg1 = trajector, arg2 = landmark
Class 2 - arg1 = landmark, arg2 = trajector
Class 3 - No relation

The classes used by the MOVELINK classifier are:

Class 1 - arg1 = mover, arg2 = goal
Class 2 - arg1 = goal, arg2 = mover
Class 3 - No relation

5.2 Evaluation

5.2.1 Setup

Once again, Stanford CoreNLP was used for POS tagging, lemmatization and dependency parsing. The classification models were trained with Vowpal Wabbit's one-against-all multi-class classifier [1] using its online stochastic gradient descent implementation with all the

Relation Type	P	R	F1
QSLINK	0.630	0.502	0.560
MOVELINK	0.529	0.533	0.531
OLINK	0.515	0.439	0.474
OVERALL	0.560	0.500	0.527

Table 4: `HRI-CRF-VW`'s relation classification results, tested on the SemEval 2015 dataset

default settings. Vowpal Wabbit uses adaptive, individual learning rates and per feature normalized updates. The initial t value is 0 with a *t power* value of 0.5.

System	P	R	F1
UTD-SPRL-SUPERVISED2	**0.610**	0.540	**0.573**
KUL-SKIP-CHAIN-CRF	0.487	0.512	0.500
UNITOR-HMM-TK	0.551	0.391	0.458
HRI-CRF-VW	0.469	**0.611**	0.531

Table 5: Relation classification results of all known SpRL systems, tested on the SemEval 2013 dataset

5.2.2 Datasets

The same dataset used for spatial element and signal detection, mentioned in Section 2.4, was also used for spatial relation classification with the exception of 9 files that didn't have spatial relations annotated. However, since our system focuses on relations with a trigger, we filtered out the relations that contained no trigger. The resulting dataset of 1,801 relations was used to train and test our system for SemEval 2015.

5.3 Results

We evaluated our system on the dataset in Section 5.2.2 using 5-fold cross validation. To get accurate performance results of the relation classification sub-system, we used gold spatial elements and signals. The results of this evaluation are shown in Table 4.

Additionally, `HRI-CRF-VW` was evaluated on SemEval 2013 Task B to compare to the other systems. However, since previous relation classification tasks were significantly different than the one proposed for SemEval 2015, we had to make a few changes to our system. We replaced the multi-class classifier with a binary classifier that simply decides whether a candidate relation tuple $<$ `TRAJECTOR`, `SPATIAL_INDICATOR`, `LANDMARK` $>$ is a valid relation. Results comparing the relation classification performance of all the systems are shown in Table 5. `UTDSpRL-SUPERVISED2` outperforms the other systems in F1 and precision, but `HRI-CRF-VW` has the highest recall.

6 Discussion

Throughout comparisons to existing systems on SemEval 2013 tasks, `HRIJP-CRF-VW` has the best recall on all tasks and the best F1 score for 2/3 of *simple roles* and 6/7 of *extended roles*. Our system also has the second highest F1 score on the relation classification task, losing only to `UTD-SPRL-SUPERVISED2`. Furthermore, despite an increase in labels and task complexity, our system has comparable performance in cross-fold validation over SemEval 2015 data.

The feature ablation results in Table 6 show the three features with the largest contribution to spatial element and signal classification. They verify the contribution of word vectors trained on Web-scale data and support `UNITOR-HMM-TK`'s [2] claims that shallow grammatical information is essential.

Evaluating our system over several iterations of the SemEval SpRL task raised several questions. First, *does splitting SpRL into spatial element/signal identification followed by role labeling make the task easier or harder?* In order to explore this, we need to determine if richer spatial element type information helps or hinders SpRL. Second, *if this SpRL task setting is indeed more difficult, how can we capture the linguistic expressiveness of its annotations while maximizing the tractability of the learning problem?* Finally, *for this new formulation, is SpRL with less (or no) feature engineering feasible?* To find out, we are exploring phrase vector-based models inspired by [4].

Features	P	R	F1	F1 Δ
all	**0.795**	**0.674**	**0.730**	-
-EF.1	0.807	0.604	0.691	-0.039
-EF.9	0.808	0.602	0.690	-0.040
-EF.10	0.761	0.600	0.671	-0.059

Table 6: The three spatial element classification features with the largest delta in feature ablation

7 Conclusion

In this paper we presented a novel system that conducts spatial role labeling using a combination of lexico-syntactic information and word vectors. Evaluation on SemEval 2013 test data showed that our system achieves a higher F1 score than all known existing systems for 2/3 of roles on a simplified spatial role identification task and all but one system on a spatial relation classification task. On a extended spatial role identification task, our system achieves a higher F1 score than the existing state-of-the-art for 6 of 7 roles. Preliminary evaluation on SemEval 2015 training data showed comparable performance despite a more difficult task setting. For future work, we are in the process of testing a phrasal-vector-based approach inspired by the SRL system of [4].

Acknowledgments

This research was supported by Honda Research Institute JP.

References

[1] Alekh Agarwal, Olivier Chapelle, Miroslav Dudík, and John Langford. A reliable effective terascale linear learning system. CoRR, abs/1110.4198, 2011.

[2] Emanuele Bastianelli, Danilo Croce, Roberto Basili, and Daniele Nardi. UNITOR-HMM-TK: Structured kernel-based learning for spatial role labeling. In Proc. of SemEval, 2013.

[3] Charles J. Fillmore, Christopher R. Johnson, and Miriam R.L. Petruck. Background to FrameNet. International Journal of Lexicography, 16.3, 2003.

[4] Karl Moritz Hermann, Dipanjan Das, Jason Weston, and Kuzman Ganchev. Semantic frame identification with distributed word representations. In Proc. of ACL, 2014.

[5] Oleksandr Kolomiyets, Parisa Kordjamshidi, Steven Bethard, and Marie-Francine Moens. SemEval-2013 task 3: Spatial role labeling. In Proc. of SemEval, 2013.

[6] Parisa Kordjamshidi, Steven Bethard, and Marie-Francine Moens. SemEval-2012 task 3: Spatial role labeling. In Proc. of SemEval, 2012.

[7] Parisa Kordjamshidi, Martijn Van Otterlo, and Marie-Francine Moens. Spatial role labeling: Towards extraction of spatial relations from natural language. TSLP, 8(3):4, 2011.

[8] Lluís Màrquez, Xavier Carreras, Kenneth C Litkowski, and Suzanne Stevenson. Semantic role labeling: an introduction to the special issue. Computational Linguistics, 34(2), 2008.

[9] Naoaki Okazaki. CRFsuite: a fast implementation of conditional random fields (CRFs), 2007.

[10] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In Proc. of EMNLP, 2014.

[11] Kirk Roberts and Sanda Harabagiu. UTD-SpRL: A joint approach to spatial role labeling. In Proc. of SemEval, 2012.