Recognizing Spatial Containment Relations between Event Mentions

Kirk Roberts
Human Language Technology Research Institute
University of Texas at Dallas
kirk@hlt.utdallas.edu

Michael A. Skinner
University of Texas Southwestern Medical Center
Children's Medical Center
michael.skinner@childrens.com

Sanda M. Harabagiu
Human Language Technology Research Institute
University of Texas at Dallas
sanda@hlt.utdallas.edu

Abstract

In this paper, we present an approach for recognizing spatial containment relations that hold between event mentions. Event mentions refer to real-world events that have spatio-temporal properties. While the temporal aspect of event relations has been well-studied, the spatial aspect has received relatively little attention. The difficulty in this task is the highly implicit nature of event locations in discourse. We present a supervised method that is designed to capture both explicit and implicit spatial relation information. Our approach outperforms the only known previous method by a 14 point increase in F_1-measure.

1 Introduction

Understanding the interaction of events expressed in natural language requires the ability to recognize spatio-temporal relations between event mentions. While the automatic recognition of temporal relations has received significant attention in the literature (Pustejovsky et al. 2003, Verhagen et al. 2009, UzZaman et al. 2012), the automatic recognition of spatial relations has received comparatively little attention. We believe this is partly due to the difficulty of the task as compared to temporal event relations. The difficulty stems from the fact that (i) spatial relations are multi-dimensional and therefore have a more complex representation, (ii) narrative structure is largely chronological in nature, and the events are often presented by their relative temporal order instead of their relative spatial arrangement, and (iii) by extension, spatial event relations are typically implicit in nature, relying on an intuitive understanding of the semantic properties of events.

Spatial relations between events that are explicitly expressed are typically indicated through syntactic relationships, such as "*The [presentation] at the [conference] was excellent*". Here, the preposition *at* indicates the *presentation* event is spatially contained within the *conference* event.

Far more common, however, are implicitly expressed spatial event relations. For example, in the sentence "*The [bombing] victim [died] immediately*", it is clear that the *died* event is spatially related to the *bombing* event. Specifically, we would say that the *bombing* event spatially contains the *died* event since the assumed bounds of the *bombing* is larger.

An automatic method for recognizing spatial relations between events would be useful for many extraction and reasoning tasks. For instance, mapping the location of entities mentioned in discourse has generally been accomplished through semantic role labeling, which links a predicate with its local semantic arguments. However, locations are relatively rare in discourse as compared to verbal and nominal predicates. Usually the location of an entity is not directly stated in the entity's local argument structure. Instead, this information is *implicit* as the relevant information is located outside a limited syntactic/semantic scope. Tying an entity to a location mentioned elsewhere in the discourse requires either co-reference (either entity or event co-reference), or an understanding of the spatial interactions present within the discourse structure so that relevant spatial inferences may be made.

The goal of this paper is to enable this type of spatial reasoning by connecting events through spatial containment relations.

These spatial relations allow for complex reasoning beyond simply placing an entity on a map. Consider the following sentence taken from a surgeon's operative note:

> A longitudinal [incision] through the umbilicus was [carried] down through to the fascia.

Here the nominalized *incision* event is spatially tied to the *carried* event. Understanding the spatial relation between these events allows us to recognize that a three-dimensional path exists from the point of the *incision* down to the *fascia*, a layer of tissue between the skin and muscles. This deep spatial understanding of text motivates new forms of information extraction, machine reading, and question answering.

In this paper, we present a mostly supervised approach to the detection of spatial event relations. Due to the presense of both explicitly and implicitly expressed relations, we rely on two different classes of features. The first class, which targets explicitly expressed relations, utilizes typical information extraction features, such as lexical and syntactic context. The second class, which targets implicitly expressed relations, focuses on identifying semantically related events that are more likely to be spatially related (such as *presentation* and *conference*, or *bombing* and *die*). This allows us to leverage unlabeled data to derive semantic similarity measures.

The remainder of this paper is organized as follows. Section 2 outlines related work in generalized event relations, generalized spatial relations, as well as current work in spatial event relations. Section 3 describes the data we use to train and evaluate our models. Section 4 details our supervised method, including our classifier, features, and feature selection technique. Section 5 contains the results of our experiments. Section 6 discusses the limitations of our approach and proposes future work. Finally, Section 7 concludes by summarizing our work.

2 Related Work

Event relations in general have received significant attention in the literature, but largely in the form of temporal event relations. The TimeML annotation standard (Pustejovsky et al., 2003) for temporal relations as well as the TimeBank corpus (Pustejovsky et al., 2003) have inspired a significant number of automatic systems for this task (Verhagen et al. 2009, Verhagen et al. 2010, UzZaman et al. 2012, Sun et al. 2013). Beyond temporal relations, work in other types of event relations has received less attention. Prominent among the other event relation types is causation (Bethard and Martin 2008, Beamer and Girju 2009, Rink et al. 2010, Do et al. 2011) and co-reference (Chen et al. 2009, Bejan and Harabagiu 2010). Beyond event relations, Chambers and Jurafsky (2008, 2009) and Bejan (2008) both create narrative schemas based on commonly co-occurring event structures, which is a useful tool for determining a prior likelihood of two or more events being related.

Spatial relations between non-events has likewise received much attention. Several such works are spatial annotation schemas. SpatialML (Mani et al., 2008) focuses on recognizing geographic regions and expressions. For example, the following text:

> a town some 50 miles south of Salzburg in the central Austrian Alps

SpatialML would recognize *town*, *Salzburg*, *Austrian*, and *Alps* as geographic locations, normalize *Salzburg* and *Austrian* to their respective geo-political entities, recognize the direction and distance relation between *town* and *Salzburg*, and the containment relations between *Salzburg* and *Austrian* and *Alps* and *Austrian*. SpatialML has no handling, however, for spatial event relations. Likewise, SpRL (Kordjamshidi et al., 2010) represents spatial relations beyond geographic relations, but would have difficulty representing event relations because SpRL requires an indicator (trigger, e.g., *in*, *on*, *at*, *to the left of*) that is rarely present in spatial event mentions. SpRL does, however, have an annotated corpus (Kordjamshidi et al., 2012) and several automatic approaches have been proposed (Kordjamshidi et al. 2011, Roberts and Harabagiu 2012). STML (Pustejovsky and Moszkowicz, 2008) focuses on the annotation of spatial relations for events, specifically motion events. But their scheme connects a motion event with its motion-specific arguments, and does not include event-event spatial relations.

Despite significant work in both event relations and spatial relations, work specific to spatial relations between events has been quite sparse. ISO-Space (Pustejovsky et al. 2011a, Pustejovsky et al. 2011b, Lee et al. 2011) is an on-going effort to develop a detailed annotation system for spatial information (beyond just spatial language). However, no publicly available corpus is known to exist.[1] Prior to this work, we have developed a corpus (Roberts et al., 2012) of spatial event relations, which is discussed in detail in the next section. While its spatial representation is not as rich as ISO-Space, it contains similar relation types and is designed to represent the highly implicit nature of spatial event relations.

3 Data

In order to conceptualize spatial relations between event mentions, the event itself must be spatially conceptualized. In Roberts et al. (2012), we suggest this can be done by approximating the spatial bounds of an event. For instance, an *election* event might assume the spatial bounds of the geo-political entity conducting the election; a sporting event may be bounded by the field or stadium in which it is played; and a *battle* event may be bounded by the immediate vicinity of the various battle participants. A spatial relation between events, then, can be determined by comparing the spatial bounds of two events, such as whether they are equal, overlap, or one event subsumes the other.

This corpus consists of 162 newswire documents, a subset of the SpatialML corpus (Mani et al., 2008). The corpus contains 5,029 events and 1,695 spatial relations. Annotators marked each event as "spatial" or not based on whether they had intuitive spatial bounds (e.g., "*the gas [attack]*" would be spatial while "*the stock price [increase]*" would not be spatial as it is not clear what the spatial bounds of *increase* might be). In order for a spatial relation to hold between two events, both events must be marked as spatial. For the purposes of this paper, we only evaluate on event pairs in which both events are manually marked as spatial. The data contains six different spatial relation types:

1. SAME: Two events E1 and E2 have indistinguishable spatial bounds.
2. CONTAINS: E1's spatial bounds contain E2's spatial bounds.
3. R_CONTAINS: E2's spatial bounds contain E1's spatial bounds.
4. OVERLAPS: E1 and E2 share partial spatial bounds but neither is a sub-set of the other.
5. NEAR: E1 and E2 do not share spatial bounds but they are within close proximity of each other.
6. DIFFERENT: E1 and E2 have distinguishably different spatial bounds.

These relation types are based on RCC-8 (Randell et al., 1992). Four of the part-of relations are collapsed into CONTAINS and R_CONTAINS. Also, NEAR and DIFFERENT replace the disconnected and externally connected relations, a design decision similar to SpatialML. An example sentence from this corpus exemplifies the CONTAINS relation:

> In October of 1985, four hijackers under his command [took] over the Italian cruise ship Achille Lauro and [killed] a wheelchair-bound American tourist, Leo Klinghoffer.

Here, the *took* event is determined to exhibit a CONTAINS relation with the *killed* event, as *took*'s spatial bounds are determined to be the entire cruise ship, while the spatial bounds of *killed* are the immediate vicinity of the victim.

In addition to annotating spatial events and spatial relations between events, the corpus contains annotated participants and locations of the events. In this way we can graphically represent the spatial relationships between various entities in the text, such as in Figure 1. This graph allows us to make the inference that *Leo Klinghoffer* was located on the *Achille Lauro* when he was killed. Without such a relation, we would have to make the (un-principled) assumption that the closest location (in this case a vehicle) is the location of the *killed* event.

4 Method

We utilize a mostly supervised, two-stage machine learning approach for detecting spatial event relations. A binary support vector machine (SVM) classifier is used for recognizing spatial relations and a multi-class SVM is used for determining the relation type. Previous SVM-based approaches to relation extraction have utilized advanced kernels (e.g., Nguyen et al. (2009)). In this work, however,

[1]Gaizauskas et al. (2012) have annotated a small corpus of facility design reports with a version of ISO-Space, but it is neither publicly available nor large enough to utilize as training data in a machine learning approach. Furthermore, the majority of its spatial relations (perhaps all) are not between events.

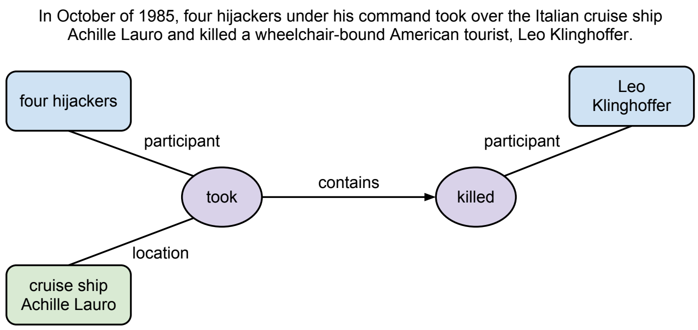

Figure 1: Example spatial event relation from our corpus.

we focus on the utility of different feature types and perform our experiments with a linear kernel using LibLinear (Fan et al., 2008). We evaluate on 3-sentence and 1-sentence windows for potentially related events (annotators were limited to relations in a 3-sentence window). Since the vast majority of event mentions are not spatially related, we adjust the negative weight to 0.05 (leaving the positive weight at 1.0). For the multi-class relation type classifier, SAME and CONTAINS make up the vast majority of relations and therefore get a weight of 0.1. These weights were tuned using a different cross-validation split than that used on our experiments below. Below we detail the features used by the two classifiers. For both classifiers, a large number of partially overlapping features were developed, most of which are described below. We utilize a greedy forward/backward technique known as floating forward feature selection (Pudil et al., 1994) to optimize the best sub-set of features.

4.1 Explicit Relation Features

These features are designed to recognize explicit statements of spatial relatedness based on the context of the relation. Sometimes explicit relations are expressed with spatial prepositions such as *in*, *on*, or *at*. In general, however, we consider explicit relations to be those in which the local context indicates a spatial relation is highly likely. For instance:

> After today's air [strikes], 13 Iraqi soldiers [abandoned] their posts and [surrendered] to Kurdish fighters.

Here, *abandoned* and *surrendered* share the same subject and are syntactically connected through the conjunction *and*. When an actor performs two actions at connected or overlapping time intervals, the actions are necessarily spatially related. While an *and* dependency doesn't necessarily guarantee temporal connectivity, it is highly suggestive and therefore acts as a good indicator.

We utilize the following classes of features:

- Words between arguments, which includes features that are ignored entirely when the arguments are separated by a certain number of tokens or sentences. These bag-of-word features provide useful lexical context that is not always available from a dependency parse (such as modifiers).
- Token-level and sentence-level distance. Event mentions that are lexically closer are more likely to be spatially related, and mentions in different sentences are much less likely to be spatially related.
- Dependency paths. We use the Stanford Dependency Parser (de Marneffe et al., 2006) with the collapsed representation so that preposition nodes (*prep*) become edges (e.g., *prep_at*). This also results in more semantic conjunction edges (*conj_and* instead of simply *conj*).
- TimeML relations from TARSQI (Verhagen et al., 2005), including TLinks, SLinks, and ALinks. TLinks are typical temporal links, such as one event occurring before or after another event. SLinks are a subordinate links, such as in "*John [promised] to [buy] wine for Mary*". ALinks are aspectual links, such as "*John [stopped] [talking]*".
- Event participants/locations from the manually annotated data. If necessary these could be automatically annotated by a semantic role labeler.

Top 5 Events via TLink PMI

bomb	PMI	**pass**	PMI	**drive**	PMI
strafe	0.298	touchdown	0.288	intoxicate	0.276
nuke	0.281	defense	0.277	floorboarding	0.271
landmark	0.273	exam	0.268	park	0.239
shell	0.242	interception	0.249	impair	0.231
machine-gun	0.242	amendment	0.233	bike	0.208

Top 5 Events via Gigaword Sentence PMI

bomb	PMI	**pass**	PMI	**drive**	PMI
strafe	0.295	touchdown	0.354	homer	0.289
plot	0.292	veto	0.237	intoxicate	0.260
nuke	0.291	vote	0.230	floorboarding	0.257
landmark	0.255	squash	0.212	reformat	0.256
scan	0.249	test	0.209	touchdown	0.234

Table 1: Highly associated events for *bomb*, *pass*, and *drive*, as acquired from unlabeled data.

4.2 Implicit Relation Features

These features are designed to recognize spatial relatedness between events based entirely on their semantic properties (i.e., without regard to context). Many times our intuitive understanding of event structures enables the omission of linguistic context clues of spatial relations. For instance:

> During a live broadcast, Geraldo [drew] a map in the sand [showing] the location of the unit in relation to Baghdad.

Here, we understand the purpose of *drew* is manifested in *showing*, and further that in such a relationship the two events are connected by a common object (in general a drawing, but specifically a *map* in this example) that forms an integral part of their spatial bounds. This kind of information requires a source of external knowledge, potentially from (i) a manually constructed knowledge base, (ii) knowledge built from training data, or (iii) knowledge built from unlabeled data. While manual knowledge sources such as ConceptNet (Liu and Singh, 2004) or FreeBase (Bollacker et al., 2008) could be utilized, they are quite sparse on event information (rather focusing on entity information). Instead, we focus on learning which individual events are likely to participate in a spatial relation (using the training data), which pairs of events are likely to participate in a spatial relation (also from the training data), and which pairs of events are likely to be related (from unlabeled data).

We utilize the following classes of features:

- Individual arguments (separate features for first and second arguments). Includes features based on event mention's surface form, caseless form, lemmatized form, part-of-speech from the Stanford Parser (Klein and Manning, 2003), General Inquirer categories (Stone et al., 1966), TimeML event classes from TARSQI (Verhagen et al., 2005), WordNet (Fellbaum, 1998) synsets and hypernyms, and VerbNet (Kipper et al., 1998) classes.
- Concatenation of the above individual argument features for both arguments (e.g., "*draw::show*" for lemmatized form, "*25.2::29.5-2*" for VerbNet classes).
- Intersection of feature values for individual argument features.
- Statistical association of events based on various resources:
  - Gigaword (Parker et al., 2009) sentence co-occurrence
  - TimeML relations on Gigaword
  - Wikipedia co-occurrence

The statistical association features discussed above are designed to elicit spatial information from data without spatial labels. To accomplish this, we start by extending the chronological narrative assumption to space. That is, the narrative not only expresses a directional path through time, but a path through space as well. Thus, events that are closer to each other in the narrative are more likely to be spatially related. The resources mentioned above are thus drawn from different representations of potential narratives. First, sentence co-occurrence in Gigaword is a means of discretizing the narrative into small, tightly related sets of events. Second, TimeML relations are designed to extract the narrative in a temporal structure. These relations have the advantage of including related cross-sentence events while excluding un-related within-sentence events. While this is more principled, TimeML is a difficult task, and any automatic technique would contain both noise and bias. Third, Wikipedia's article structure is more inclined to articles whose events take place in a single location. Thus, we can relax our local constraint to allow for document-wide context. This not only reduces sparsity, but is more likely to capture transitive spatial relations.

While all of these resources should be capable of providing related events, we require a method to increase the likelihood of the event associations being spatial. For this purpose, we use the statistical association metric known as pointwise mutual information (PMI):

$$PMI(x,y) = log\frac{p(x|y)}{p(y)}$$

Where co-occurrences with less than 10 instances are discarded. PMI is a simple technique that has been shown to be effective at natural language tasks, most appropriately narrative chain construction (Chambers and Jurafsky, 2008). Due to the large amount of data, we require a highly efficient technique, such as PMI, that only requires a limited view of the data.[2]

The result of these PMI calculations for three events (*bomb*, *pass*, and *drive*) are shown in Table 1. As can be seen, PMI across this data is able to capture spatially related events: *bomb* is spatially related to sub-types of bombing events such *nuke* and *shell* and other war activities such as *strafe* and *machine-gun*. PMI captures spatially related events for multiple senses of *pass* and *drive*. For instance, *touchdown*, *defense*, and *interception* are spatially related events to the sporting sense of *pass*, while *vote*, *veto* and *amendment* are spatially related events to the political sense of *pass* (as in, "*pass a bill into law*"). Further, as can be seen in Table 1, while the different data sources assign different weights, there is some degree of overlap between them.

Given the different data sources, and the myriad of potential features that could be written to represent this data (in addition to all the other feature types), we utilize the automated feature selection technique discussed above. This enables us to optimize how we present these partially overlapping features to the classifier, ultimately resulting in increased performance. We next discuss the actual features chosen by this technique.

4.3 Selected Features

The features chosen by the feature selector for relation detection are shown in Table 2. The feature selector chose four explicit relation features and eight implicit relation features. The chosen implicit relation features include the first feature chosen and five of the first six features. The features chosen by the feature selector for relation type classification are shown in Table 3. Here, the feature selector chose only two features, both of which are implicit features, suggesting the context is of little significance for determining specifically how two event mentions may be related. The next section evaluates these two classifiers on held-out data.

5 Experiments

We experiment under two different settings: (1) intra-sentence relations only, and (2) intra-sentence relations up to a 3-sentence window, the maximum relation length for the data. We evaluate both relation recognition (whether two event mentions have a spatial relation between them) and relation type classification (given a related pair of mentions, which is the proper relation type). These are both evaluated on the data described in Section 3. In Roberts et al. (2012), we present a baseline method for both spatial relation recognition and spatial relation type classification based on the event mention words, the words between the mentions, and the mention hypernyms. We consider this our baseline for the task. The results for spatial relation recognition are shown in Table 4, and the results for spatial relation classification are shown in Table 5.

Our method easily outperforms the baseline for spatial relation recognition with a 30% increase in F_1-measure. The overall score is still quite low, however, owing to the difficulty of the task. This is discussed more in the next section. Spatial relation type classification outperforms the baseline,

[2]For instance, our same sentence data has 837 million event pairs (14 million unique), while our TLink data has 360 million event pairs (12 million unique).

#	Type	Feature Description
1[a]	I	Concatenated event mention lemmas. Argument order is ignored by representing lemmas in orthographic order. E.g., *kill::take*
2	E	Dependency path between the event mentions. E.g., ↓*conj_and*
3	I	TLink co-occurrence from Gigaword, adjusted by point-wise mutual information (PMI). Specifically, we use a symmetric PMI so the feature is mention-order independent. This is done by taking the minimum of PMI(E1, E2) and PMI(E2, E1). (real-valued)
4	I	Concatenated event mentions in their caseless form. Argument order is preserved. E.g., *took::killed*
5	I	Co-occurrence from Wikipedia, adjusted using PMI. (real-valued)
6	I	Concatenated event mention lemmas. Argument order is preserved unlike Feature 1. E.g., *take::kill*
7	E	Whether the two event mentions have the same location. This feature uses the Stanford co-reference resolution system (Raghunathan et al., 2010) to expand locations so that two events have the same location if their respective locations belong to the same co-reference chain. (boolean-valued)
8	E	Whether the two event mentions have the same participant. (boolean-valued)
9	E	Token distance between the event mentions. Reduced to scalar between 0 and 1 by computing $1 - (\lvert t_1 - t_2 \rvert + 1)^{-1}$. (real-valued)
10	I	Intersection of event mention categories from the General Inquirer. E.g., *kill*'s categories are: ACTIVE, DAV, H4LVD, HOSTILE, NEGATIV, NGTV, NOUN, PFREQ, SOCREL, STRONG, SUPV, and TRNLOSS. *take*'s categories are: ACTIVE, AFFIL, BEGIN, DAV, FETCH, H4, HANDELS, IAV, MODIF, NEED, POWER, SOCREL, STRONG, SUPV, TRY, VARY, and VIRTUE. The intersection is thus ACTIVE (i.e., active orientation), DAV (descriptive action verb), SOCREL (socially defined inter-personal process), STRONG (strength), and SUPV (support verb).
11	I	Intersection of event mention VerbNet classes. E.g., $\emptyset$
12	I	Concatenated event mention surface form. Argument order is ignored. E.g., *killed::took*

Table 2: Spatial event relation recognition features, shown in the order chosen by the feature selector. Type 'E' refers to the explicit features (Section 4.1), Type 'I' refers to the implicit features (Section 4.2). Feature values taken from example in Figure 1.

#	Type	Feature Description
1	I	Whether the ALink co-occurrence PMI (from Gigaword) is greater than 0 (i.e., is the aspectual link positively correlated?). This does not use a symmetric PMI because the relation type order matters. (boolean-valued)
2	I	Co-occurrence from Gigaword sentences, adjusted using PMI. (real-valued)

Table 3: Spatial event relation type features.

[a]This was not technically the first feature chosen. Instead, the length of the dependency path was the first feature, but this was pruned after Feature 8 was added to the feature set.

Method	1-sentence P	R	F_1	3-sentence P	R	F_1
Baseline	35.1	41.3	37.9	29.1	35.5	32.0
Our Method	44.7	69.2	54.3	37.2	60.4	46.0

Table 4: Spatial event relation recognition experiments on our corpus.

Method	1-sentence %	3-sentence %
Baseline	59.3	58.3
Our Method	60.1	59.3

Table 5: Spatial event relation type classification experiments on our corpus.

but only slightly. Here, the issue is largely a matter of data imbalance: the SAME relation is favored by the classifier in almost all cases.

Reducing the context to a single-sentence window improves the relation recognition score by a further 8.3 points. While this would limit the reasoning power of any downstream system, it is useful to know that performance gains are possible by focusing on an easier sub-set of the data. This improvement in relation recognition does not apply to relation type classification, however. In the next section we place our results in greater context and analyze some typical errors.

6 Discussion

The performance gains seen in the previous section are encouraging: they validate our assumption that spatial information can be obtained from large amounts of unlabeled data in an efficient manner. The overall F_1-measure, though, still seems quite low compared to other natural language tasks such as named entity recognition (NER) and semantic role labeling (SRL). However, those tasks are limited to explicit context, such as contiguous tokens for NER and parse nodes within the syntactic scope for SRL. These tasks also utilize more predictable features, such as surface-level casing features for NER and predictable argument structures for SRL (e.g., the syntactic subject for an active verb is usually the ARG0). Proper comparison requires evaluating our results alongside other implicit tasks. One such work involves implicit SRL. Gerber and Chai (2010) perform nominal SRL and achieve an overall F_1-measure of 42.3. While the tasks are not directly comparable in terms of difficulty, this does suggest that implicit tasks require far more advanced methods to achieve superior performance and that downstream systems will likely need to be highly tolerant to noise. To address this, we discuss future work below, analyzing the types of errors that our system makes to give context to these ideas.

As might be guessed, rare event mentions with long dependency paths are highly likely to result in false negatives, such as the relation between *elected* and *disillusionment* here:

> Tehran had been governed by reformists since 1989, but a conservative city council was [elected] in the February 28 municipal polls in a result attributed to a meager turnout amid growing public [disillusionment] with electoral politics.

Here the *elected* and *disillusionment* events are judged to cover all of Tehran. The dependency path for this relation has five edges, including the rare dependency relation *prep_amid*. Further, the *disillusionment* event is fairly rare. Such long dependency relations with rare arguments is unlikely to be recognized by a simple machine learning classifier. Instead, this suggests an approach where either intermediate events are able to transitively suggest spatial relations, or the dependency parse is relaxed in certain cases to allow for longer-range relations.

As is common in semantic tasks, word sense presents an issue, resulting in a false negative:

> It was believed Naotia was a [practicing] sorcerer and through his black magic he had [cast] evil spells on villagers, prompting a group within the village to eliminate them.

Since our corpus-based method uses a lemmatized form only, when related but rare senses are used, such as the witchcraft sense of *cast*, PMI is unable to attribute the proper association between the two events.

In terms of false positives, our implicit features can result in errors when very similar events are clearly different based on their context:

> The British leader [travelled] to the United States before also [visiting] Japan, South Korea, China on a whistle-stop tour.

Here, the spatial bounds of *travelled* is interpreted as being the United States and the flight from Britain, while the *visiting* event is interpreted as being several Asian countries. While one might argue these two trips are spatially related since one is a continuation of the other, the annotator in this case chose to use neither the NEAR or OVERLAPS relations. This highlights another issue with such implicit tasks: the annotations rely heavily on the annotator's intuition. Not unexpectedly, the corpus has fairly low inter-annotator agreement (Roberts et al., 2012).

One final error highlights the difference between events that are related by a narrative, and events that are spatially related:

> Police have [arrested] four people in connection with the [killings].

This false positive resulted from the high degree of association between *arrested* and *killings*, but arrests are rarely made at the scene of the crime. One potential solution to this is to automatically extract event narrative structures, then check the locations of the events on that structure for unexpected location changes. This would be quite challenging: automatic narrative structures proposed thus far are quite simplistic, and most events within a narrative structure will not have an explicit location, so a very robust model of structure would be required.

Finally, despite the accuracy score being higher than the F_1-measure for relation recognition, spatial relation type classification may be the more difficult task. Almost all errors were the result of misclassifying a relation as SAME due to the class imbalance. While the classifier weights may be tuned to improve F_1-measure for recognition, this rarely improves a multi-class task significantly. Our main direction for future work is to actually classify the *size* of events. For example, we would like to recognize that an *election* has larger bounds than a *protest*. This would allow our classifier to recognize when two events are very different in size, and if so which is larger. Ideally, by constricting the set of classes for a containment relation using the sizes of the arguments, this would allow other semantic features to contribute to relation type classification.

7 Conclusion

We have presented an approach for recognizing spatial containment relations between event mentions. Using a corpus of event mentions from newswire texts, we have developed a supervised classifiers for (1) recognizing the presense of a spatial relation between two event mentions, and (2) classifying spatially related event pairs into one of five spatial containment relations. Our method combines features that are designed to extract explicit information from the relation context, as well as implicit information about the likelihood of two events being spatially related. We have evaluated our method and shown substantial improvements over the pre-existing baseline, achieving an F_1 of 46.0 on relation recognition and an accuracy of 59.3% on relation type classification. These gains, though, are largely limited to the task of recognizing whether a spatial relation exists. Finally, we have performed an error analysis to determine paths of future work on this challenging task.

References

Beamer, B. and R. Girju (2009). Using a Bigram Event Model to Predict Causal Potential. In *Proceedings of Computational Linguistics and Intelligent Text Processing*, pp. 430–441.

Bejan, C. A. (2008). Unsupervised Discovery of Event Scenarios from Texts. In *Proceedings of the 21st Florida Artificial Intelligence Research Society International Conference (FLAIRS)*.

Bejan, C. A. and S. Harabagiu (2010). Unsupervised Event Coreference Resolution with Rich Linguistic Features. In *Proceedings of the Association for Computational Linguistics*.

Bethard, S. and J. H. Martin (2008). Learning Semantic Links from a Corpus of Parallel Temporal and Causal Relations. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 177–180.

Bollacker, K., C. Evans, P. Paritosh, T. Sturge, and J. Taylor (2008). Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge. In *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 1247–1250.

Chambers, N. and D. Jurafsky (2008). Unsupervised Learning of Narrative Schemas and their Participants. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.

Chambers, N. and D. Jurafsky (2009). Unsupervised Learning of Narrative Schemas and their Participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*.

Chen, Z., H. Ji, and R. Haralick (2009). A pairwise event coreference model, feature impact and evaluation for event coreference resolution. In *Proceedings of the RANLP Workshop on Events in Emerging Text Types*, pp. 17–22.

de Marneffe, M.-C., B. MacCartney, and C. Manning (2006). Generating Typed Dependency Parses from Phrase Structure Parses. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*.

Do, Q. X., Y. S. Chan, and D. Roth (2011). Minimally Supervised Event Causality Identification. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 294–303.

Fan, R.-E., K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin (2008). LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research 9*, 1871–1874.

Fellbaum, C. (Ed.) (1998). *WordNet: An Electronic Lexical Database*. MIT Press.

Gaizauskas, R., E. Barker, C.-L. Chang, L. Derczynski, M. Phiri, and C. Peng (2012). Applying ISO-Space to Healthcare Facility Design Evaluation Reports. In *Seventh Workshop on Interoperable Semantic Annotation (ISA), Eighth International Conference on Language Resources and Evaluation*, pp. 13–20.

Gerber, M. and J. Y. Chai (2010). Beyond NomBank: A Study of Implicit Arguments for Nominal Predicates. In *Proceedings of the Association for Computational Linguistics*, pp. 1583–1592.

Kipper, K., H. T. Dang, and M. Palmer (1998). Class-based construction of a verb lexicon. In *Proceedings of AAAI/IAAI*.

Klein, D. and C. D. Manning (2003). Accurate Unlexicalized Parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pp. 423–430.

Kordjamshidi, P., S. Bethard, and M.-F. Moens (2012). SemEval-2012 Task 3: Spatial Role Labeling. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval)*.

Kordjamshidi, P., M. V. Otterlo, and M.-F. Moens (2011). Spatial Role Labeling: Towards Extraction of Spatial Relations from Natural Language. *ACM Transactions on Speech and Language Processing 8*(3).

Kordjamshidi, P., M. van Otterlo, and M.-F. Moens (2010). Spatial role labeling: Task definition and annotation scheme. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, pp. 413–420.

Lee, K., A. C. Fang, and J. Pustejovsky (2011). Multilingual Verification of the Annotation Scheme ISO-Space. In *International Conference on Semantic Computing (ICSC)*, pp. 449–458.

Liu, H. and P. Singh (2004). ConceptNet – a practical commonsense reasoning tool-kit. *BT Technology Journal 22*(4), 211–226.

Mani, I., J. Hitzeman, J. Richer, D. Harris, R. Quimby, and B. Wellner (2008). SpatialML: Annotation Scheme, Corpora, and Tools. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*.

Nguyen, T.-V. T., A. Moschitti, and G. Riccardi (2009). Convolution kernels on constituent, dependency and sequential structures for relation extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 1378–1387.

Parker, R., D. Graff, J. Kong, K. Chen, and K. Maeda (2009). English Gigaword Fourth Edition. *The LDC Corpus Catalog. LDC2009T13*.

Pudil, P., J. Novovičová, and J. Kittler (1994). Floating search methods in feature selection. *Pattern Recognition Letters 15*, 1119–1125.

Pustejovsky, J., J. Castano, R. Ingria, R. Saurí, R. Gaizauskas, A. Setzer, G. Katz, and D. Radev (2003). TimeML: Robust Specification of Event and Temporal Expressions in Text. In *IWCS-5 Fifth International Workshop on Computational Semantics*.

Pustejovsky, J., P. Hanks, R. Saurí, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo (2003). The TIMEBANK Corpus. In *Proceedings of Corpus Linguistics*.

Pustejovsky, J. and J. L. Moszkowicz (2008). Integrating Motion Predicate Classes with Spatial and Temporal Annotations. In *Proceedings of COLING 2008*, pp. 95–98.

Pustejovsky, J., J. L. Moszkowicz, and M. Verhagen (2011a). Iso-space: The annotation of spatial information in language. In *Proceedings of the Sixth Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*, pp. 1–9.

Pustejovsky, J., J. L. Moszkowicz, and M. Verhagen (2011b). Using ISO-Space for Annotating Spatial Information. In *Proceedings of the International Conference on Spatial Information Theory*.

Raghunathan, K., H. Lee, S. Rangarajan, N. Chambers, M. Surdeanu, D. Jurafsky, and C. Manning (2010). A Multi-Pass Sieve for Coreference Resolution. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*.

Randell, D. A., Z. Cui, and A. G. Cohn (1992). A Spatial Logic based on Regions and Connection. In *Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning*, Volume 117.

Rink, B., C. A. Bejan, and S. Harabagiu (2010). Learning Textual Graph Patterns to Detect Causal Event Relations. In *Proceedings of the 23rd Florida Artificial Intelligence Research Society International Conference (FLAIRS'10), Applied Natural Language Processing Track*, pp. 265–270.

Roberts, K., T. Goodwin, and S. M. Harabagiu (2012). Annotating Spatial Containment Relations Between Events. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pp. 3052–3059.

Roberts, K. and S. M. Harabagiu (2012). UTD-SpRL: A Joint Approach to Spatial Role Labeling. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval)*.

Stone, P. J., D. C. Dunphy, and M. S. Smith (1966). *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

Sun, W., A. Rumshisky, and O. Uzuner (2013). Evaluating Temporal Relations in Clinical Text: 2012 i2b2 Challenge Overview. *Journal of the American Medical Informatics Association Submitted*.

UzZaman, N., H. Llorens, J. F. Allen, L. Derczynski, M. Verhagen, and J. Pustejovsky (2012). TempEval-3: Evaluating Events, Time Expressions, and Temporal Relations. *arXiv.1206.5333v1*.

Verhagen, M., R. Gaizauskas, F. Schilder, M. Hepple, J. Moszkowicz, and J. Pustejovsky (2009). The TempEval challenge: identifying temporal relations in text. *Language Resources and Evaluation 43*, 161–179.

Verhagen, M., I. Mani, R. Saurí, R. Knippen, J. Littman, and J. Pustejovsky (2005). Automating Temporal Annotation with TARSQI. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Demo Session*.

Verhagen, M., R. Saurí, T. Caselli, and J. Pustejovsky (2010). SemEval-2010 Task 13: TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 57–62.