Enhancing Spatial Relationships Detection in text: Analyzing methodologies and managing ambiguity

OLEKSII KYRYK, University of Twente, The Netherlands

ABSTRACT

Spatial relationship detection in natural language poses a complex challenge due to the inherent ambiguity and nuances of human language. This paper explores the methodologies for detecting spatial relationships in natural language, considering how the various approaches deal with inherent ambiguity and implicit spatial information in text. The study compares past methodologies, introduces an experiment of its own, and discusses strategies employed to manage ambiguity in spatial relationship detection.

1 INTRODUCTION

Humans use language to describe the world around them. Vital part of using language is talking about physical objects and the relationship between them in some physical space. A sentence "The laptop is on the table and the mouse is to the left of it." describes the relationship between 3 entities – the laptop, the table, and the mouse. To be more specific, it explains the position (or region) of the laptop relative to the table as well as the direction (or orientation) of the mouse with respect to the laptop. Detecting such spatial arrangements and learning the mapping onto a formal representation poses significant difficulties due to the inherent ambiguity of human language.

Detecting spatial relationships within text holds substantial relevance across several domains, including robotics, traffic management, and image/graphics generation (Tappan, 2004). Spatial knowledge, however, is often implicit in natural language. This is one of the biggest challenges in enabling natural communication between people and intelligent systems (Chang, Savva, & Manning, 2014). For instance, if we want a robot to retrieve a mouse given the example sentence above, it needs to be inferred that the mouse is to the left of the laptop, even though it is not explicitly stated. In the scene generation application, using the same sentence, the scene generation software must have an understanding of the likely location of the table relative to the space it is in (the table location is not mentioned in the sentence).

This research paper will focus on exploring and investigating a range of approaches and methodologies used in the spatial relationships detection task. By analyzing the strengths and weaknesses of various techniques, especially in contexts where spatial relationships are implicit or indirectly expressed, this research aims to contribute to managing the inherent ambiguity of human language in spatial relationship detection.

TScIT 40, February 2, 2024, Enschede, The Netherlands

2 PROBLEM STATEMENT

Detecting spatial relationships in natural language presents complex challenges that become significantly more difficult due to the inherent ambiguity and implicit spatial cues within human discourse. Research shows that extracting spatial entities and their relationships from text can be achieved with high accuracy in simple sentences that include an explicit spatial indicator (a token that defines constraints on the spatial properties, e.g. *in, on, behind*) (Kordjamshidi, Van Otterlo, & Moens, Spatial role labeling: Towards extraction of spatial relations from natural language, 2011), however, the task becomes notably more complex when dealing with multifaceted descriptions. Ambiguity is particularly found in sentences that describe more than one relationship. In such cases finding the correct, or the intended link between the objects described in text becomes a lot more difficult. Another example is sentences that utilize a verb instead of a preposition as a spatial indicator. These sentences usually require some "common sense" to determine whether the sentence describes a physical spatial relationship. For example, in the sentence "He left the room 5 minutes ago", the sentence semantics indicate that "He" is not in the "room" which is a spatial relationship between "He" and "room" A different sentence, "He left the church 10 years ago" describes the person's faith – a fundamental change and not the spatial relationship between "He" and "church". Both sentences use the same verb "left" but have different semantic meanings. Such sentences, that use motion indicators (Zlatev, 2006) are significantly more complex for spatial relationships detection task. Disambiguating such examples for spatial role labeling remains challenging, especially when spatial elements and their relationships span multiple sentences.

This research has 2 objectives. The first objective is to analyze how various methodologies and techniques were used in the past to achieve spatial relationships detection in text. Spatial relationship detection has been an active topic of research in recent years, with researchers performing experiments by employing different methods on datasets of different languages. Some spatial role-labeling models were designed on a context-specific dataset of a certain language while others had an objective to achieve context and language-independent solutions. By looking into and contrasting these methodologies, this research aims to uncover nuances of spatial relationship detection, considering the challenges of different linguistic structures and contexts. From exploring several different methodologies and approaches in spatial relationship detection across various datasets, this research will dive deeper into the management of ambiguity inherent in natural language and implicit spatial information. The research will discuss how different methods tackle ambiguity as well as what are the remaining challenges in managing ambiguity and implicit spatial information.

2.1 Research questions

To address the challenges and goals outlined in the problem statement, this study will answer the following research questions:

- **RQ1:** How do the existing approaches and methodologies for spatial relationship detection in natural language compare?
- **RQ2:** How can natural language processing models effectively manage ambiguity and implicit spatial information to enhance the accuracy and contextual understanding of spatial relationships in text?

3 RELATED WORK

The spatial relationships detection problem in text is a relatively novel problem that became a subject of research in the early 2000s, In the research of this topic, application-dependent relations could be extracted from text in specific languages (Kelleher, 2003) (Tappan, 2004).

In their 2010 paper, Kordjamshidi et al. (Spatial Role Labeling: Task Definition and Annotation Scheme.), have introduced the spatial role labeling problem as the extraction of generic spatial semantics from natural language and in their 2011 article (Spatial role labeling: Towards extraction of spatial relations from natural language), they tackled the problem of spatial role extraction from unrestricted natural language with machine learning methods. In this article, however, they have restricted their focus on prepositions only and have focused on machine learning techniques to achieve their goals. While taking a huge step forward in tackling the spatial role labeling problem in this research, the restriction of using sentences with prepositions only means that a lot of ambiguous sentences (such as ones involving motion indicators) were not used for spatial relationships detection.

There has been research investigating spatial relations detection task from different angles, such as rule-based (Zhang, Zhang, & Jiang, 2010), and machine learning (Kordjamshidi, Van Otterlo, & Moens, Spatial role labeling: Towards extraction of spatial relations from natural language, 2011), (Wu & Zhang, 2023). The paper by Wu & Zhang similarly to my research objectives, aimed to compare different approaches to spatial relationship detection. Their investigation compared the pipeline and joint extraction approaches.

In the following sections of this paper, you will find an in-depth analysis of the methods and results of each of the aforementioned papers to create a meaningful comparison between different approaches to spatial relationship detection.

Managing ambiguity is quite a broad topic with many different examples in human language, and there has been different research tackling different examples of the problem. For example, in the previously mentioned 2011 article, Kordjamshidi et al. (Spatial role labeling: Towards extraction of spatial relations from natural language), use the research by Litkowski and Hargraves (SemEval-2007 Task 06: Word-Sense Disambiguation of Prepositions) for preposition disambiguation.

There has been research on managing ambiguity by resolving prepositional phrase attachments through visually guided spatial relationship detection – a multimodal approach taking both text and image for spatial relationship detection (Rahgooy, Manzoor, & Kordjamshidi, 2018).

4 DATASETS

This section serves as a vital introduction to addressing research question 1, aiming to compare different methodologies employed in past spatial relationship detection research. Understanding the nature and diversity of the datasets is an essential part of analyzing the diverse methodologies and techniques in spatial relationship detection research. This section describes the datasets utilized to evaluate each methodology under comparison within this paper.

4.1 Machine Learning Approach

In this research, I will examine several papers proposing solutions to spatial relationship detection via a machine-learning approach. Hence there is an extensive list of datasets to consider while evaluating different results. Here is the list of datasets that were employed in the research and their characteristics:

a) *TPP dataset* (Kordjamshidi, Van Otterlo, & Moens, Spatial role labeling: Towards extraction of spatial relations from natural language, 2011)

TPP dataset was used for the preposition disambiguation task. It contains 34 XML files, totaling 16557 training and 8096 test example sentences in English, with each sentence containing one example of a preposition.

b) *GUM (Maptask) dataset* (Kordjamshidi, Van Otterlo, & Moens, Spatial role labeling: Towards extraction of spatial relations from natural language, 2011)

A subset of the *GUM (Maptask) dataset* was used for the spatial role labeling task. 100 English sentences were used in this dataset, with 65 trajectors and 69 landmarks appearing in 112 spatial relations.

c) *CLEF dataset.* (Kordjamshidi, Van Otterlo, & Moens, Spatial role labeling: Towards extraction of spatial relations from natural language, 2011)

This dataset of textual descriptions of 400 images contains 686 English sentences that contain 839 trajectors and 741 landmarks. The total number of spatial relations in this dataset is 869.

d) *DCP dataset* (Kordjamshidi, Van Otterlo, & Moens, Spatial role labeling: Towards extraction of spatial relations from natural language, 2011), (Kolomiyets, Kordjamshidi, Moens, & Bethard, 2013)

This dataset contains descriptions of locations situated at each of the latitude and longitude integer degree intersections in the world. Not all sentences are in English, unlike the previously mentioned datasets. In (SemEval-2013 Task 3: Spatial Role Labeling, 2013), 1789 sentences from the corpus (in original orthography and formatting) were used with 2198 trajectors and 1353 landmarks appearing in 2703 spatial relationships. In (Kordjamshidi, Van Otterlo, & Moens, Spatial role labeling: Towards extraction of spatial relations from natural language, 2011), however, only 250 sentences were used with 199 trajectors and 188 landmarks appearing in 222 spatial relationships. This dataset is particularly interesting to consider as it is collected user data and hence a large proportion of the sentences involving a preposition did not describe a spatial relationship.

e) *IAPR TC-12 image Benchmark dataset* (Kordjamshidi, Bethard, & Moens, SemEval-2012 Task 3: Spatial Role Labeling, 2012)

This dataset contains images taken by tourists with text descriptions in different languages. This dataset consists of 1213 sentences, in which 1593 trajectors and 1408 landmarks appear in 1464 spatial relationships.

6.1.1 Multi-modal datasets.

As mentioned in the previous section, one of the approaches to managing ambiguity is using multimodal data for training. Therefore, the following datasets were considered:

f) *CLEF 2017 mSpRL dataset* (Rahgooy, Manzoor, & Kordjamshidi, 2018)

This dataset is a subset of the previously mentioned *IAPR TC-12 image Benchmark dataset* which contains 613 images with descriptions of 1213 sentences.

g) *Visual Genome dataset (VG)* (Rahgooy, Manzoor, & Kordjamshidi, 2018)

The dataset contains 108077 images and 2316104 relation instances (the relationships component of this dataset contains relationships (prepositions) between two bounding boxes).

h) *ReferItGame dataset* (Rahgooy, Manzoor, & Kordjamshidi, 2018)

It contains 120000 expressions which cover 99.5% of the regions of the *SAIAPRTC-12 dataset* which is an annotated version of the *IAPR TC-12 image Benchmark dataset.*

4.2 Chinese Datasets

The Encyclopedia of China (the geography section) has been used as a dataset in several papers researching spatial relationship detection from China. The rule-based approach (Zhang, Zhang, & Jiang, 2010) has utilized this dataset for their research. It is important to note that it is difficult to draw conclusions from the results of Chinese and English datasets due to linguistic nuances and structural differences between the languages.

4.3 Experiment Dataset

Summarizing the comprehensive overview of all the datasets (and their respective papers), it is noteworthy to highlight that I am going to use a subset of the *IAPR TC-12 image Benchmark dataset* when conducting my own experiment. The choice to use a subset of the *IAPR TC-12 image Benchmark dataset* for my experiment is motivated by its common use across several of the methodologies discussed in this paper, therefore allowing a more meaningful analysis and evaluation. Section 6.2 will describe the experiment in further detail.

5 EVALUATION

In answering Research Question 1 concerning the analysis of various methodologies and approaches in spatial relationship detection, a series of evaluation metrics must be defined to discuss their relative strengths and weaknesses. The comparison will be conducted based on the following criteria to provide a comprehensive understanding of each approach's performance across different datasets.

5.1 Performance Metrics

The first and one of the most important criteria for evaluation is of course comparing the performance metrics across different models within the Machine Learning approach to the spatial relationship detection task. These metrics will include precision, recall, and F1 score as they are the standard evaluation metrics employed by all the papers mentioned earlier.

5.2 Dataset-specific Analysis

As mentioned in the Datasets section, there are a lot of datasets to consider in this research. Recognizing the influence of the dataset, its characteristics, and preparation, on a model/methodology performance is vital to compare it with alternatives. To add to that, recognizing the dataset contents will help draw conclusions about how the methodology deals with ambiguity of natural language. Specific attention will be given to instances where the same dataset was used in different approaches – that gives a great chance to have a fair comparison of their relative performance.

5.3 Contextual considerations

Lastly, recognizing the diverse application of spatial relationships detection, an important consideration is the context of the research behind the solution proposed. Some approaches may have fine-tuned their models to a specific context, such as the Geographic Information Systems (GIS) (Zhang, Zhang, & Jiang, 2010), while others have pursued a context-independent solution. In the evaluation of the results, the implications of the contextual choices on the methodologies' performance must be considered.

6 METHODOLOGIES

This section will delve into the comprehensive methodologies employed in spatial relationship detection. Understanding the different choices made in previous research will help draw conclusions for cross-methodology comparison. This section will describe the nuances and processes of various machine learning approaches to the problem, uncovering the differences across multiple techniques. Additionally, it will illustrate how other non-machine learning techniques contribute to the spatial relationship detection problem. Moreover, this section will introduce the discussion of how different methodologies handle ambiguity, which will be continued in the Results section. Lastly, in this section, you will find a detailed report of my experiment and how it differs from the rest of the methodologies considered in this research.

6.1 Machine Learning Approach

In the field of spatial relationship detection within natural language, the machine learning approach has emerged as the most common method for solving the problem. Numerous studies have employed the machine learning approach; however, I took the work of Kordjamshidi et al. across several papers and conferences as it is the most comprehensive solution proposed in recent years. Furthermore, analyzing the solutions presented in their collective body of work holds significance, given the interconnected nature of their papers. The *SemEval* series forms a continuum, building upon prior contributions. This sequential progression allows for a nuanced examination of the evolution of techniques, pinpointing the variations instrumental in enhancing the outcomes and expanding the scope of the spatial relationship detection problem. This section will describe the machine learning solutions to the spatial relationships detection, outlining their differences, followed by a section describing my experiment, employing different choices in my approach to the ones employed in past research. All the machine learning methodologies described in this section aimed at context-independent models and were trained (mostly) on datasets in the English language. The comparison of the results within different machine learning models will be in the Results section.

One of the most foundational works of research on spatial relationship extraction from the Machine Learning perspective is the paper (Spatial role labeling: Towards extraction of spatial relations from natural language) by (Kordjamshidi, Van Otterlo, & Moens) in 2011. This work, introducing the spatial role labeling task, delves deeply into machine learning methodology for spatial relationship detection, addressing ambiguity and implicit spatial information inherent in human discourse.

The spatial role labeling task aims to assign the words in a given sentence one of the elements of the set of roles: trajector, landmark, spatial_indicator, or none. In this paper, **only prepositions** were considered as spatial indicators though. While the authors acknowledge that various word classes can act as a spatial indicator in a sentence, they argue that the most dominant form is the preposition. Therefore, another arguably common type of spatial pivot - a verb, was not considered in this research. The sentences in the datasets used by the researchers containing such examples of spatial indicators were omitted in data pre-processing. Hoverer, while such an example of the ambiguity of language was not addressed in this research, the paper did address a different problem - the implicit landmark or trajector. One of the most common kinds of ambiguity is the non-explicit subject of the spatial relationship, as in the example given in the introduction of this paper (the is implicit in the relation *toTheLeftOf(mouse, laptop)).* The approach of the authors of the (Spatial role labeling: Towards extraction of spatial relations from natural language) suggests another term "undefined": to highlight the existence of implicit trajectors or landmarks in a sentence. Consequently, the authors have defined two steps for spatial role labeling: 1) the function that takes a word in the sentence as an input and classifies whether this word is a spatial indicator. They employ a probabilistic classifier for this. 2) Given the fact that some word is a spatial indicator, a multi-class classifier is employed to tag the rest of the words of the sentence into roles {trajector, landmark, none}. After the words in a sentence have been labeled with their roles corresponding to some spatial pivot, no learning is required to produce the spatial triplets. The results of the spatial elements extraction and the generated triplets will be compared with the following methodologies in the Results section. The performance of spatial role extraction (given a spatial pivot) will be evaluated in 2 settings: 1) Where the spatial pivot is known (ground truth) and 2) where the spatial indicator classification is used as input for trajector/landmark extraction.

Based on the work of their previous paper that introduced the task of spatial role labeling, in their next work, (SemEval-2012 Task 3: Spatial Role Labeling), the authors focused on predicting the existence of spatial information at the sentence level as well as classifying the type of relation which can be *direction, region* or *distance.*

The SemEval-2012 task was defined in three parts. 1) The extraction of the individual roles (spatial indicator, trajector, and landmark) from the input sentence. 2) The extraction of the spatial triples and 3) The classification of the type of the spatial relation. As this approach to the spatial relationship detection follows from the previous work on the topic, the authors tackled implicit trajectors or landmarks with a special term "undefined". As you notice this paper is essentially an extension of the previous work, with the first 2 parts of the task being the same as the ones in (Spatial role labeling: Towards extraction of spatial relations from natural language). However, the methodology differs in producing the spatial relationships triplets. To produce spatial roles for the candidate triplets (spatial indicator, trajector, landmark) a joint classification was used, with 2 different runs which differed in the number of features. Then the binary support vector machine classifier was employed to predict whether a candidate triple is a spatial relationship or not, while in the previous work, since the extraction of trajector and landmark was based on the spatial pivot as input, the triplets were produced without learning and were not verified. After the spatial triplet has been verified, the type of the relationship is classified between the tree types mentioned above. Both runs (with larger and smaller numbers of features) will be compared to each other as well as to the other machine learning methodologies in the Results section.

The second iteration of the task (Kolomiyets, Kordjamshidi, Moens, & Bethard, 2013) extends the previous work, by using an additional training corpus, which besides containing "static" spatial relationships (identified by a preposition), also contains motions.

In the previous works, methodologies of which were previously described, the only spatial relationships considered were those between still objects. In addition to those, (SemEval-2013 Task 3: Spatial Role Labeling) paper introduced new spatial roles to capture dynamic relationships: such relationships contain an object whose location is changing in a sentence describing it. In this new annotation scheme, the roles are slightly different. Trajector of a dynamic spatial relationship denotes an object which moves in a sentence. A *motion indicator* (replacing the spatial indicator in a static relationship) is a word or a phrase that signals the motion of the trajector, which is usually a verb. A *path* is a role assigned to a word or a phrase that denotes a path of the motion that the trajector is moving along. A landmark is a secondary (and static) object of the spatial scene. Lastly, *distance* and *direction* have different meanings from the previous annotation scheme. In the 2012 paper, distance and direction were types of spatial relationships (which were classified for the triplet in the third part of the task). Here, *distance* is a role assigned to a word in a sentence if a sentence describes a motion that is performed for a certain distance. Similarly, *direction* is assigned to a word that mentions the direction of motion in the text. The goal of the SpRL-2013 can be summarised with the following 5 parts: 1) Identifying the words or phrases for three types of spatial roles such as trajector, landmark, and spatial indicator. Essentially, this part is the same as before as just static relationships would be considered here 2) Identifying the triplets that connect the roles identified in part 1. 3) Identifying the words or phrases for **all** spatial roles such as trajector, landmark, spatial indicator, motion indicator, path, direction and distance. 4) Identifying *n-tuples* that connect the roles identified in part 3. 5)

Semantic classification of the spatial relationships. In tasks 1 and 3, each word in a sentence is classified with respect to the possible spatial roles. To avoid overfitting on the training data, another vital contribution of this methodology was the use of shallow grammatical features instead of the full syntax of the sentence. Like the previous *SpRL* papers, this methodology first classifies spatial and motion indicators and then uses these predictions for the remaining spatial roles to be classified. For classifying spatial and motion indicators, lexical features such as lemmas, part-of-speech tags, and lexical context representations were used by the classifier. For task 2, similarly to the (SemEval-2012 Task 3: Spatial Role Labeling) paper discussed in this section, a support vector machine classifier would verify whether candidate triplets are spatial relationship triplets.

6.2 My Experiment

As an extension of my exploration into diverse machine learning approaches for spatial relationship detection, this section describes my experiment which was aimed to use the insights from the previous research, however, differs in approach from some of the methodologies described in the section above.

Before I dive into describing (in more detail) the dataset used for my experiment and the model, to be consistent with the rest of the approaches' descriptions I have summarized my experiment's goal in the following 3 parts. 1) Assign a spatial label (spatial_indicator, trajector, landmark, or none) to each word in the sentence **simultaneously**. (Spatial Role Labelling) 2) Generate candidate spatial relationships triplets and 3) Verify if the candidate triplet is indeed a spatial relationship. The general approach is quite similar to (a combination) of approaches that I have previously discussed. However, part 1 differs in the fact that I label the sentence words without taking a spatial pivot into account. The reasoning behind this choice is to investigate how labeling the words simultaneously differs in terms of performance.

6.2.1 Dataset description and preprocessing

As mentioned previously, for my experiment I have used the same dataset that was used in a number of approaches, for instance in (Kordjamshidi, Bethard, & Moens, SemEval-2012 Task 3: Spatial Role Labeling, 2012). Before describing the experiment, I will describe the dataset in more detail, as well as how I chose to preprocess it. The IAPR TC-12 dataset is an XML file of the following structure:

```
<TEXT>
  The entire text corpus
</TEXT>

<TAGS>
        <TRAJECTOR         text="text"         end="end_index"
        start="start_index" id="id" />

        <LANDMARK          text="text"         end="end_index"
        start="start_index" id="id" />

        <SPATIAL_INDICATOR     text="text"     end="end_index"
        start="start_index" id="id" />

        <RELATION                        general_type="region"
        spatial_indicator_id="S_id"
        landmark_id="L_id"trajector_id="T_id" id="id" />
</TAGS>
```

The <TEXT> field contains a total of 599 sentences which contains 716 trajectors and 661 landmarks that a total of 765 spatial relationships consist of. As you may have noticed, the dataset I am using for my experiment is a *subset* of the dataset that was used in (Kordjamshidi, Bethard, & Moens, SemEval-2012 Task 3: Spatial Role Labeling, 2012), which had almost double the amount of sentences.

To achieve part 1, the dataset needed to be preprocessed to create a train and test sets for spatial role labeling. The preprocessing for part 1 resulted in the dataset of the following structure:

```
<SENTENCES>
  <SENTENCE>
         <TEXT />
         <LABELS />
  </SENTENCE>
</SENTENCES>
```

The <TEXT> field of the original dataset was used together with the <TAGS> to produce a dataset where for each <SENTENCE> in <SENTENCES> there are 2 subfields: <TEXT> and <LABELS> which would contain the human text (tokenized on the word level and labels for each word. Here is an example entry from this dataset:

```
<SENTENCE>
         <TEXT>
                  ['a', 'rock', 'hole', 'in', 'the', 'ground',
'surrounded', 'by', 'sand', '.']
         </TEXT>

         <LABELS>
                  ['TRAJECTOR',    'TRAJECTOR',    'TRAJECTOR',
'SPATIAL_INDICATOR',          'LANDMARK',           'LANDMARK',
'SPATIAL_INDICATOR', 'SPATIAL_INDICATOR', 'LANDMARK', 'NONE']
         </LABELS>
</SENTENCE>
```

All sentences were padded and words were converted to numerical values for the model to use. 479 sentences would be used for training (80% train test split).

6.2.2 Model

Initially, I started my experiment with the intent of using the BERT (Bidirectional Encoder Representations from Transformers) model which is renowned for its contextual understanding of natural language. To add to that, I have seen limited research on the machine learning approaches using BERT for spatial role labeling task. However, during the experimentation phase, the training time required for BERT and the attained poor results motivated me to switch gears and explore simpler models that would be more advantageous for a small dataset such as the one I am using. Therefore, a different method was a sequential model that uses LSTM layers. The model was provided an encoded input text and would predict a spatial role for each element of the input sequence. After the fine-tuned model reached acceptable results, the output was used for parts 2 and 3. Generating candidate spatial relationship triplets required no learning similar to (Spatial role labeling: Towards extraction of spatial relations from natural language). For step 3, a binary classifier was used to verify if a candidate triplet is a spatial relationship or not. The approach was to train the binary classifier on the ground truth data (the true triplets from the dataset) and then use the trained classifier on the generated triplets in part 2.

The results of my experiment will be discussed and compared to the rest of the approaches in the Results section.

6.3 Rule-based approach

The work of (Zhang, Zhang, & Jiang, 2010) proposed a rule-based approach to the spatial relationship detection. Besides a completely different approach, this work also differs in both language and context. The dataset employed for this research is the Chinese encyclopedia (Geography section) as the original data, obviously in the Chinese language. The context of the research was the Intelligent geographical information systems, therefore the spatial relation corpus mainly consisted of relationships between named geographical entities. Based on the dataset, more than fifty syntactic patterns were identified. After these have been identified, the GATE platform was used as the natural language processing tool to implement the automatic rule-based extraction of spatial relationships. JAPE (a Java Annotation Patterns Engine) which is a module of GATE provides the ability to create the rules based on the patterns that were identified. Based on that, a natural language rule library was built. Before the experiment could be run, some modifications to the GATE platform were introduced due to a lack of support for Chinese word processing. 1) A new plugin was developed and added to GATE which segmented and POS-tagged Chinese text. 2) The annotations of geographical named entities and spatial relationships were defined in GATE.

7 RESULTS & DISCUSSION

In this segment, the outcomes of my research will be discussed. The performance across different methodologies and approaches will be evaluated based on the evaluation metrics defined earlier. Besides evaluating the performance of the models and comparing the results with dataset and context considerations, this section will compare how each approach manages ambiguity, addressing research question 2.

7.1 Ambiguity

Before diving into comparing the performance of different approaches and methodologies, this sub-section will reflect on how different methodologies manage ambiguity.

7.1.1 Spatial Indicator Ambiguity

When dealing with static spatial relationships that use a preposition as a spatial pivot (which has been the case for most of the machine learning approaches discussed in this paper), the prepositions' sense can be ambiguous. From my experiment, when analyzing the dataset, I encountered the following sentence: "About 20 kids in traditional clothing and hats waiting on stairs.". In this sentence, there is an "in" preposition which has no spatial meaning, and an "on" proposition which does. Such examples are very common in human language for static as well as for dynamic relationships that usually have a verb as a motion indicator (remember the example in the introduction, "He left the room 5 minutes ago" vs. "He left the church 10 years ago"). The (Spatial role labeling: Towards extraction of spatial relations from natural language) paper has addressed this by performing preposition disambiguation as "step 0" in their approach. They have used Naïve Bayes and Max-Entropy classifiers for the detection of spatial vs non-spatial preposition senses. They conducted the experiment by training the classifiers on the TPP dataset and compared the results with the (Litkowski & Hargraves, 2007). The results in Table I show that the proposed solution to preposition disambiguation significantly outperforms the results of (SemEval-2007 Task 06: Word-Sense Disambiguation of Prepositions) and conclude that these high-frequency prepositions can be reliably disambiguated, and hence can be identified correctly as spatial pivots in a sentence. The researchers implemented 34 classifiers for the prepositions based on the evaluation results. For some prepositions (for example "opposite"), no classifier existed at the time of the research. Therefore, the researchers had to omit the examples where "opposite" was in a sentence – as there was no reliable way to identify if the sentence containing the "opposite" preposition had spatial sense. Such practice of removing examples to avoid ambiguity turned out to be a common approach to ambiguity management for spatial relationship detection.

Table I. Results of Preposition Disambiguation on frequently used prepositions in the TPP Dataset

Preposition	Naïve Bayes			MaxEnt			SRL (2007)		
	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1
on	0.733	0.963	0.832	0.788	0.950	0.861	0.707	0.399	0.510
after	0.500	0.900	0.643	0.540	0.700	0.609	0.000	0.000	0.000
in	0.660	0.920	0.769	0.697	0.882	0.779	0.558	0.906	0.691
before	0.670	0.857	0.750	0.800	0.570	0.666	0.500	0.428	0.461

In my own experiment, part 1 of my approach was to label the spatial elements simultaneously, including the spatial pivot (preposition). For such common prepositions as the ones in Table I, the dataset used for my experiment contained examples with prepositions of both spatial and non-spatial meanings. Therefore, for such prepositions, there was ambiguity in some sentences of the dataset (such as the example given above). Since the train set of my experiment provided to the model is a sentence and corresponding labels, the train set contains examples of both spatial and non-spatial prepositions with correct labels. Therefore, my assumption is that the model learns to differentiate them in the learning process. After looking at the results, of spatial labeling of sentences (part 1) that contained common prepositions (such as those in Table I), I can conclude that the model labels the sentences with good performance. Table II shows the precision, recall, and F1 score of the spatial role labeling model for sentences with "on", and "in" prepositions. As the dataset used for my experiment was limited, there were 0 examples of sentences with "after" prepositions and only 1 example of a sentence with a "before" preposition. Mind that these results should not be directly compared with those in Table 1 as the preposition disambiguation classifies the preposition as being spatial or non-spatial while my spatial role labeling model labels all words in a sentence with their spatial roles. Nevertheless, the results in Table II demonstrate that with enough examples in the training set, the preposition can be disambiguated by the model with a simultaneous spatial role labeling approach.

Table II. Results of Spatial Role Labelling of sentences with "on" and "in" prepositions

Preposition	num_of_sentences	Pre	Rec	F1
on	112	0.849	0.837	0.839
in	353	0.934	0.930	0.931

Consequently, it is important to mention that the approach of labeling all words simultaneously suffers at labelling sentences which have a preposition that is uncommon in the training data. The approach taken by Kordjamshidi et al. across several papers of classifying the spatial indicator first and then using that as input for extracting its corresponding trajector(s) and landmark(s) performs better for low-frequency spatial indicators.

7.1.2 Motion Indicator Ambiguity

As described in the Methodology section, in (SemEval-2013 Task 3: Spatial Role Labeling), the authors have considered dynamic spatial relationships as well as static. Therefore, besides dealing with spatial pivots' ambiguity, motion indicators (verbs) can have spatial and non-spatial meanings as well. As the motion indicators are classified first before the rest of the new spatial roles are assigned, by looking at the results of the motion indicator classification, we can make conclusion on how this methodology managed motion indicator ambiguity. The precision, recall and F1 score for the motion indicator classification was 0.892, 0.294 and 0.443 respectively. The F1 score already suggests that recognizing the motion indicator is no "easy" task, compared to classifying a spatial indicator with an F1-score of 0.926 reported by the authors after evaluating part 1 of their experiment (classification of static roles: trajector, landmark and spatial indicator). The low recall of 0.294 may suggest that the model incorrectly classifies a motion indicator as non-spatial, meaning a high number of False Negatives. These results of motion indicator classification suggests that ambiguity of human language is difficult to manage in context of dynamic spatial relationship detection.

7.1.3 Ambiguity management in Rule-based approach

Thus far, this subsection has discussed how different machine learning methodologies address management of ambiguity within the spatial relationship detection in text. However, as my research extends to understanding how ambiguity is managed across different approaches, it is important to analyze the results of the rule-based approach.

The rule-based approach has not gathered as much attention in the spatial relationships detection research community as could be deduced from a brief description of this methodology in the section above. Notably, the sole paper that proposed rule-based solution differs drastically from the machine learning methodologies analyzed in this research. Besides the (Rule-Based Extraction of Spatial Relations in Natural Language Text) paper aiming to extract spatial relationships in Chinese language (resulting in potentially significantly different rules due to linguistic structural differences to English), the authors also specifically targeted to solve the problem within GIS domain. This makes ambiguity discussion for the rule-based approach less relevant, as due to the nature of the context of the dataset, there are no examples of ambiguous sentences, unlike context-independent sentences present in a variety of datasets employed by the machine learning approaches. As there is a lack of context-free rule-based solutions in the field of spatial relationships detection, it is challenging to draw conclusions to the effectiveness of this method's ambiguity management. However, I expect that ambiguity would be a great problem in context-free rule-based spatial relationship detection. The root of the challenge is because rule-based approaches determine spatial relationships by comparing the input sentences against defined rules, which, in turn rely on part-of-speech tags of the input, which is exactly where most of the ambiguity inherent in human language occurs.

7.1.4 Ambiguity due to multiple relationships in a sentence

Another sentence encountered in my dataset was "there are red umbrellas in a park on the right.". There are clearly 2 spatial relationships described in the sentence. However, it is unclear what are the 2 spatial triplets that should be accepted in part 3. One spatial relationship is very clear - *in(red umbrellas, park)*. However, the last part of the sentence creates ambiguity. Is the spatial relationship *onTheRight(red umbrellas, park)*? Or is it *onTheRight(park, undefined)* - is the park a trajector instead of a landmark with the landmark being implicit (on the right from the perspective of the person viewing the park). The true labels of spatial roles of the sentences have park labeled as a landmark and hence that implies that the correct 2nd spatial relationship is *onTheRight(red umbrellas, park)* (which is verified by looking at the `<RELATION>` tag in the dataset). By looking at the predictions, the model correctly simultaneously predicts both spatial indicators of the sentence ("in" and "on the right") as well as the trajector - "red umbrellas". However, due to this ambiguity, it fails to label the "park" with neither landmark nor trajector, instead labelling it with none. Because of this, the outcomes of part 2 will result in incorrect triplets which in turn will impact the verification in part 3.

One approach suggested in the (Visually Guided Spatial Relation Extraction from Text), aims to resolve such examples, where there are multiple spatial relationships causing ambiguity, by exploiting the help of the associated image. The image would be used to find the right relationship in such cases. While the results of evaluating that approach did show that using associated images resulted in improved ambiguity resolution, the consequence of needing an image as well as text as input to the model raises practical concerns to this solution.

7.1.5 Ambiguity - conclusion

In conclusion of this subsection discussing research question 2, the exploration of diverse machine learning methodologies shows that there are common cases of ambiguity that can be effectively managed. Spatial indicator can be determined quite well with multiple approaches able to disambiguate prepositions. Dynamic relationships are more challenging as motion indicators (which are usually a verb) are evidentially way "harder" to determine to have spatial or non-spatial meaning.

The story of managing ambiguity is a story of managing expectations. Realistically, "universal ambiguity management" is not feasible, as due to linguistic intricacies of human language. In the example in section 7.1.4, both interpretations of the ambiguous second spatial relationship are semantically valid, however only one option is the ground truth in the dataset. The pragmatic truth to ambiguity management is that while common instances can be feasibly addressed, some cases simply need to be overlooked. That can include removal of edge cases from the dataset (something that has been done by the approaches I analyzed), or more domain specific resolution. For example, if a spatial relationships detection model is utilized for 3D scene generation, and an ambiguous sentence is given as input, for such example the model may return all semantically valid options (such as the 2 options of the example above) back to the user in the form of a generated 3D scene, and a final user input is required to resolve the ambiguity.

7.2 Performance Analysis

This sub-section will discuss and compare the performance of the diverse approaches and methodologies discussed in this paper. The performance will be analyzed based on the performance metrics stated in the Evaluation section.

7.2.1 Spatial Role Labelling

While methodologies differed slightly, the general first step of each machine learning approach was assigning spatial roles to the elements of the input sentence. Table III lists the precision, recall, and F1 score of each methodology's results discussed in this paper, including my experiment. For the (Spatial role labeling: Towards extraction of spatial relations from natural language, 2011) paper, 2 settings are compared (rows 2 and 3), using ground truth prepositions and the classified prepositions as input. The performance metrics for the spatial indicator were not reported by the authors for row 3 which is why they are missing from this table. From (SemEval-2012 Task 3: Spatial Role Labeling, 2012) research, 2 runs (rows 4 and 5) are compared which differed in the number of inputs (larger and smaller number of inputs respectively. Lastly, the results of the approach employed in (SemEval-2013 Task 3: Spatial Role Labeling, 2013) are used in the comparison, with UNITOR-1.1 using the IAPR TC-12 image Benchmark dataset and UNITOR-2.1 stating results using the Conference Project dataset (which contained the dynamic relations as well as static). Important to note that my experiment, approach taken in 2012 runs and UNITOR-1.1 (2013) use the same dataset, which makes the comparison of the results more meaningful. The results of the 2011 methodology have used a different dataset for training which explains surprisingly much better results than the following SemEval papers.

Table III. Results of Spatial Role Labelling of different methodologies

Method	Trajector			Landmark			Spatial Indicator		
	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1
My experiment	0.940	0.876	0.907	0.395	0.939	0.556	0.905	0.859	0.881
Pip-ground truth (2011)	1.000	0.969	0.983	0.947	1.000	0.971	N/A	N/A	N/A
Pip-PP-input (2011)	0.934	0.945	0.936	0.720	0.760	0.727	-	-	-
UTDSpRL-1 (2012)	0.731	0.621	0.672	0.871	0.645	0.741	0.928	0.712	0.806
UTDSpRL-2 (2012)	0.782	0.646	0.707	0.894	0.680	0.772	0.940	0.732	0.823
UNITOR-1.1 (2013)	0.684	0.681	0.682	0.741	0.835	0.785	0.967	0.889	0.926
UNITOR-2.1 (2013)	0.565	0.317	0.406	0.661	0.476	0.554	0.612	0.481	0.538

As mentioned in the Methodologies section, the SpRL papers analyzed in my research have been employing a different approach to the one I have in my experiment. Looking at the results, the classification of the spatial indicator first and then extraction of its corresponding landmark(s) and trajector(s) has a significant advantage for classifying the Landmark class. That is where my proposed methodology evidently suffers with low precision implying many false positives. However, Trajector is classified significantly better in my simultaneous approach in comparison to the other methodologies that use the same dataset for training. In addition, the contribution of (SemEval-2013 Task 3: Spatial Role Labeling) paper - the use of lexical features such as lemmas and part-of-speech tags in classifying the spatial indicator gives significant improvement (as seen in UNITOR-1.1 row) relative to all other approaches.

7.2.3 Triplets generation performance

All approaches' next step is the generation of triplets. The performance of this approach is summarized in Table IV which includes the results of the Rule-based approach as well. The performance metrics are based on the binary classifier that verifies the generated triplets.

Table IV. Results of Spatial Triplets generation across Machine Learning approaches and the Rule-based approach

Method	Precision	Recall	F1-Score
My experiment	0.470	0.680	0.550
UTDSpRL-1 (2012)	0.567	0.500	0.531
UTDSpRL-2 (2012)	0.610	0.540	0.573
UNITOR-1.1 (2013)	0.431	0.306	0.358
Rule-based approach (2010)	0.742	0.654	0.694

The results demonstrate that the generated triplets' are "difficult" to verify. Looking at the table above, some might assume that the rule-based approach suddenly is a clear dominant way of extracting spatial relationships from natural text. However, the reason for the high performance of the model is because of a very specific context, and hence cannot be compared to the machine learning approaches above. Thus, a valid conclusion can be made that for niche domains, a rule-based approach to relationship detection can be advantageous.

7.2.2 Dynamic Relationships

Lastly, the Dynamic spatial relationships detection results of (SemEval-2013 Task 3: Spatial Role Labeling) will be discussed. UNITOR-2.1 model was used for part 3 of the experiment and has not reached desirable results as of the time of the research. With F1 scores of 0.443, 0.427, 0.264, and 0.490 for motion indicator, path, direction, and distance labels respectively, it is clear that more complex text in the data leads to rather poor model performance. To add to that, as seen in the last row of Table III, the classification of the static roles (Trajector, Landmark, and Spatial Indicator) in the dataset which includes both static and dynamic relations, has produced poor results.

8 CONCLUSION

In this paper, a thorough comparison of spatial relationship detection approaches and methodologies was conducted. The results of the investigation show the impact that differences between approaches have had on the performance and how the context of the data can affect the choice of methodology for spatial relationship detection. A detailed discussion of how different approaches deal with ambiguity reveals that many cases can be dealt with in a variety of techniques. However, due to the nature of human language, some edge cases will always remain.

REFERENCES

[1] Tappan, D. A. (2004, May). Knowledge-based spatial reasoning for automated scene generation from text descriptions. Las Cruces, New Mexico, USA.

[2] Chang, A., Savva, M., & Manning, C. (2014). Learning Spatial Knowledge for Text to 3D Scene Generation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 2028–2038). Doha: Association for Computational Linguistics.

[3] Kordjamshidi, P., Van Otterlo, M., & Moens, M.-F. (2011). Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Trans. Speech Lang. Process., 8*(3), 1-36.

[4] Zlatev, J. (2006). Semantics of Spatial Expressions. In K. Brown, *Encyclopedia of Language & Linguistics*.

[5] Kelleher, J. D. (2003, September). A perceptually based computational framework for the interpretation of spatial language. Dublin, Ireland.

[6] Kordjamshidi, P., Van Otterlo, M., & Moens, M.-F. (2010). Spatial Role Labeling: Task Definition and Annotation Scheme. *Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta.

[7] Zhang, C., Zhang, X., & Jiang, W. (2010). Rule-Based Extraction of Spatial Relations in Natural Language Text. *2009 International Conference on Computational Intelligence and Software Engineering* (pp. 1-4). IEEE.

[8] Wu, K., & Zhang, X. (2023, Septembed). Deep learning models for spatial relation extraction in text. *Geo-spatial Information Science, 26*(1), 58-70.

[9] Litkowski, K. C., & Hargraves, O. (2007). SemEval-2007 Task 06: Word-Sense Disambiguation of Prepositions. *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)* (pp. 24-29). Prague: Association for Computational Linguistics.

[10] Rahgooy, T., Manzoor, U., & Kordjamshidi, P. (2018). Visually Guided Spatial Relation Extraction from Text. In M. Walker, H. Ji, & S. Amanda (Ed.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers),. 2*, pp. 788-794. New Orleans: Association for Computational Linguistics.

[11] Kolomiyets, O., Kordjamshidi, P., Moens, M.-F., & Bethard, S. (2013). SemEval-2013 Task 3: Spatial Role Labeling. In S. Manandhar, & D. Yuret (Ed.), *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). 2*, pp. 255-262. Atlanta: Association for Computational Linguistics.

[12] Kordjamshidi, P., Bethard, S., & Moens, M.-F. (2012). SemEval-2012 Task 3: Spatial Role Labeling. *The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012). 1*, pp. 365-373. Montréal: Association for Computational Linguistics.