A Simple Neural Approach to Spatial Role Labelling

by

Nitin Ramrakhiyan, Girish Palshikar, Vasudeva Varma

in

European Conference on Information Retrieval

Report No: IIIT/TR/2020/-1

Centre for Language Technologies Research Centre
International Institute of Information Technology
Hyderabad - 500 032, INDIA
April 2020

A Simple Neural Approach to Spatial Role Labelling

Nitin Ramrakhiyani[1,2]✉, Girish Palshikar[1], and Vasudeva Varma[2]

[1] TCS Research, Tata Consultancy Services Ltd., Pune, India
{nitin.ramrakhiyani, gk.palshikar}@tcs.com
[2] International Institute of Information Technology, Hyderabad, India
vv@iiit.ac.in

Abstract. Spatial Role Labelling involves identification of text segments which emit spatial semantics such as describing an object of interest, a reference point or the object's relative position with the reference. Tasks in SemEval exercises of 2012 and 2013 propose problems and datasets for Spatial Role Labelling. In this paper, we propose a simple two-step neural network based approach to identify static spatial relations along with the three primary roles - Trajector, Landmark and Spatial Indicator. Our approach outperforms the task submission results and other state-of-the-art results on these datasets. We also include a discussion on the explainability of our model.

Keywords: Spatial Role Labelling, Spatial Representation and Reasoning, Deep Learning, BiLSTM

1 Introduction

Spatial Role Labelling (SpRL) is the process of assigning segments of text in a sentence, with roles they perform based on their spatial semantics. In natural language sentences describing spatial information, there is generally an object whose spatial position is being described (the *Trajector* role), a reference object (*Landmark*) and a spatial trigger (*Spatial Indicator*). There are other roles like *Path* and *Motion Indicator* which describe the dynamic position of a *Trajector*. SpRL is similar to Semantic Role Labelling (SRL) on certain counts and dissimilar on various others. It is similar to SRL mainly because both consider a central element whose arguments have to be found. Spatial indicators and motion indicators in SpRL are like verbs in SRL and other roles like Trajector, Landmark and Path are the arguments of these indicators. SpRL is however different from SRL as the central element may not always emit a spatial sense or otherwise can be part of several spatial relations.

Tasks on SpRL were introduced as Task 3 at SemEval 2012 [5], as Task 3 at SemEval 2013 [4] and as Task 8 (SpaceEval) at SemEval 2015 [10]. The tasks saw a moderate participation with organizers also providing baseline systems in some cases. The tasks introduced the various spatial roles and their semantics, while increasing the complexity of the problem each year. In Task 3 at SemEval

2012, the core task of spatial role labelling was introduced involving identification of roles namely *Trajector*, *Landmark*, *Spatial Indicator* and static relations among these roles. In Task 3 at SemEval 2013, apart from the previous year's problem, the task involving identification of dynamic relations was added. In the SpaceEval 2015 task, identification of finer roles along with their attributes was introduced.

In this paper, we focus only on the identification of static spatial relations and the roles *Trajector*, *Landmark* and *Spatial Indicator*. Hence, we do not attempt the dynamic spatial relation identification sub-problem of Task 3 at SemEval 2013. Also, we do not tackle any problems of SpaceEval 2015 due to introduction of new notion of spatial entities, change in relations to MOVELINK, QSLINK and OLINK and change in evaluation of the relation identification sub-problems.

We propose a simple two step neural approach for these tasks. We train a BiLSTM for a sequence labelling task of identifying spatial roles only to develop context vectors for the words. We then use contexts from this pre-trained BiLSTM for a relation classification step and deduce the corresponding roles from identified relations. The proposed neural model outperforms the participating systems and other state-of-the-art approaches on the datasets of the two tasks. As part of the analysis, we also discuss on the semantics of the context embeddings learned by the BiLSTM.

2 Relevant Literature

2.1 SemEval 2012 Task 3

Task 3 at SemEval 2012 [5], introduced the basic task of spatial role labelling which involved two sub-problems: identification of the three roles namely *Trajector*, *Landmark*, *Spatial Indicator* and identification and classification of static spatial relations involving these roles. The task data was a subset of image descriptions available as a part of the IAPR TC-12 image benchmark [3]. The image descriptions described entities in the images and their relative or absolute positions with respect to other entities in the image. As per the task, each spatial relation is formed of a *Trajector*, a *Spatial Indicator* and an optional *Landmark* and the relation type is classified as: *region* (describing topology such as `on`, `inside`, etc.), *direction* (describing orientation such as `above`, `to the left of`) and *distal* (describing distance such as `far`, `away`, etc.).

As an example from the dataset, consider the sentence: `a woman and a child are walking over the square.` Here, positions of the `woman` and `child` are being described and hence are the *Trajectors*. The `square` is the reference entity working as a *Landmark* and `over` describes the position of the Trajectors with respect to the Landmark, thus being the *Spatial Indicator*. There are two relations which can be identified: (`woman`, `over`, `square`) and (`child`, `over`, `square`). Both relations are *direction* type of relations.

2.2 SemEval 2013 Task 3

Task 3 at SemEval 2013 extended the task introduced in SemEval 2012 in two directions. Firstly, they considered full phrase spans of the text identified under a role instead of only head words as considered in the SemEval 2012 task. Secondly, they introduced a sub-problem of identification of dynamic spatial relations and corresponding roles like *Path*, *Direction* and *Distance*. The change from head words to full spans prompted the organizers to change from the annotation format used in SemEval 2012 to a different character based offset format.

2.3 Other Work

Another pivotal task - Task 8 (SpaceEval) at SemEval 2015 extended the earlier tasks in multiple directions. The organizers followed the annotation scheme specified in the ISOSpace standard [11] to enhance the granularity of the spatial semantics and used a more fine grained set of annotation tags. The sub-problems introduced covered identification of spatial entities which may be Places or Objects, identification of static and dynamic relations and identification of various attributes of these relations.

Mazalov et al. [8] proposes a CNN based technique similar to a one used for semantic role labelling for the tasks in Spatial Role Labelling. The authors report results on static spatial relation extraction on the SemEval 2013 dataset. Kordjamshidi and Moens [6] propose a structured learning based solution for spatial ontology population from text and report results on the datasets of SemEval 2012 and SemEval 2013 tasks. In another work [7], the authors propose visually informed embedding of words (VIEW) for use in a spatial arrangement prediction task. The paper reports its results on the SemEval 2013 Task 3 dataset.

3 Proposed Approach

We hypothesize that the context of spatial text elements is useful in finding their spatial roles. Also, employing their word level features along with their context can boost this classification further. We propose a two step approach to achieve spatial relation identification and deduction of spatial roles of the involved elements.

As the first step, we train a BiLSTM to get context embeddings for words of the sentences in the dataset. The BiLSTM is trained for a sequence labelling task of identifying spatial indicators and spatial actors. For this step, we denote both trajectors and landmarks as a single type of entity namely "spatial actor". As the only goal of this step is to learn the context of the words it is not necessary to learn a sequence labeller for the three roles separately. Moreover, it is observed that some words play multiple roles in a sentence i.e. a trajector in one relation and landmark in another. For example, in the sentence `a man sitting on a bench in front of the wall.`, the word `bench` behaves as a landmark in the relation (`man`, `on`, `bench`) but behaves as trajector in the relation (`bench`, `in` `front of`, `benchwall`). This is observed for more than 10% sentences in the datasets which restricts us from posing it as a sequence labelling problem over all the three roles. The output of this network is not utilized in the later step of the approach. This network is formed of an input layer feeding into a BiLSTM layer followed finally by a prediction layer for each time step.

As part of the second step, we first generate a list of candidate relations from each sentence. We develop a dependency parsing based candidate generation logic by analysing the dependency parse of multiple sentences in training data along with their spatial roles. The candidate generation procedure checks for each preposition in the sentence and marks its prepositional child as a possible landmark and its head as a possible trajector, if they are nouns. If they aren't nouns the algorithm continues traversing the dependency path further on each side till a noun is found and is marked as trajector or landmark. If a trajector found is connected to another noun by a conjunction dependency relation that noun is also added as a possible trajector. Landmarks are also expanded similarly. The candidate generation procedure returns a list of relation triples of trajector, spatial indicator and landmark for each sentence.

We then add the class information to each candidate relation. If the candidate relation is present in the true relations for the sentence, the relation's type (region, direction or distal) is added as its class. Otherwise the relation is assigned a null type. A second neural network is developed and trained for the task of classifying these candidate relations. The network has a input layer with three input vectors each corresponding to trajector, spatial indicator and landmark, followed by a hidden layer and finally a prediction layer. As input to the network, we propose two configurations - (i) only the context of trajector, spatial indicator and landmark of the candidate relation or (ii) context concatenated with the original (not retrained) word embeddings of the trajector, spatial indicator and landmark. The context of each word is obtained by passing the sentence through the pre-trained BiLSTM of the first step and collecting the hidden layer output at each word from both directions.

While testing, each sentence gets its candidate relations generated. For each candidate relation, context of its trajector, landmark and spatial indicator is obtained from the BiLSTM trained on training data. The candidate relation is input to the trained relation identification network and if predicted as true the corresponding trajector, landmark and spatial indicator are also marked as true. The network output and deduced roles are processed to generate annotations as per the evaluation scripts provided by the organizers.

Changes for Task 3 of SemEval 2013 For this task, it was required to predict the full span of the spatial role and not just the head word as in the earlier task. The training and testing learned from the earlier task are re-used as the data remains the same and annotations change marginally. While making predictions on the testing data the network predicts only head words as part of spatial roles. These predicted head words are expanded using another dependency based procedure. For each trajector and landmark all their determiner, compound, adjective modifier, numerical modifier and adverb modifier children are included to form the complete span. For spatial indicators, constructions of the form "on the left of" are developed starting from the predicted indicator "on".

4 Experimentation, Evaluation and Analysis

Experimentation: We use keras [2] to implement the neural networks. We use 300 dimensional Glove word embeddings [9] trained on a common web crawl of 42 billion words. For arriving at the right neural network parameters, we use five fold cross validation on the training data. For the BiLSTM network, we find the best results for 10 epochs, batch size of 32, dropout of 0.1 and 300 LSTM units. For the relation identification network, the best parameters are 10 epochs, batch size of 32, dropout of 0.3 and 600 hidden units for the context based network and 900 hidden units for the context + embeddings based network. To account for randomness in network weight initialization, we carry out the training (and testing) 10 times and report averaged results over the runs.

Evaluation: We use the evaluation scripts provided by the task organizers as part of the released dataset, to compute the results. We produce the output of our approaches as desired by the evaluation jar files, run the jar files and report results thus obtained. This puts our approach at par in comparison to the participating systems and other state-of-the-art.

For Task 3 at SemEval 2012, along with the baseline results provided by the organizers, we use the results reported in the best run [12] submitted by the single participating team, as a baseline for comparison. The results from the best approach proposed in Kordjamshidi and Moens [6] on the IAPR TC-12 dataset is also included as a baseline for comparison. In the interest of space, we request readers to refer to [6] for details on the approaches.

For Task 3 at SemEval 2013, we use the results reported in the best run [12] submitted by the single participating team, as a baseline for comparison. We only report results under the relaxed evaluation criteria as specified by the organizers. We also compare with the best results proposed in [7]. We however, do not compare with the approach in Mazalov et al. [8] as it is not clear from the paper whether the authors evaluate using the organizer provided scripts. The authors mention changing the format of the data leading to the possibility of final evaluation being carried out differently.

Table 1. F1 scores of various systems for the SemEval 2012 tasks

Approach	**TR**	**LM**	**SP**	**Relation**	**Relation Type**
Organizer Baseline [5]	0.646	0.756	**0.900**	0.500	NA
UTD Best Run [12]	0.707	0.772	0.823	0.573	0.566
EtoE-IBT-CLCP [6]	0.673	0.797	0.869	0.617	NA
Context	0.835	0.856	0.883	0.775	0.706
Context + Embeddings	**0.848**	**0.875**	**0.900**	**0.794**	**0.741**

Table 2. F1 scores of various systems for the SemEval 2013 static spatial relation tasks

Approach	TR	LM	SP	Relation
UNITOR Best Run [1]	0.682	0.785	**0.926**	0.458
VIEW [7]	0.732	0.678	0.749	0.235
Context	**0.823**	**0.814**	0.901	**0.562**
Context + Embeddings	0.808	0.8	0.878	0.556

Analysis: The pre-trained BiLSTM when tested using five fold cross validation on the training data, showed superior results of F1 greater than 0.9 for both classes - spatial indicator and spatial actor.

It can be observed from Table 1 and Table 2 that our hypothesis stating use of context and embeddings for text elements to predict spatial roles gets established. For Task 3 of SemEval 2012, our approaches outperform all baselines by substantial margins. It is important to note here that the UTD submission [12] relies on a fixed list of prepositions as spatial indicators which though a curated list, can be limiting in many cases. Our relation candidate generation logic does not rely on a fixed list and considers each preposition for a possible relation. Also the context and embedding approach is seen to perform better than the only context based approach. For the SemEval 2013 task, our context based approach shows improvement in relation identification performance over the baselines. Analysis reveals that the change from head words to full span of the roles has lead to certain inconsistencies in the annotations. This is also highlighted by authors in [8]. The lower relation identification performance on this task can be attributed to these changes.

To understand the semantics captured in the context of spatial indicators in the BiLSTM network we perform an experiment. We check whether the context of spatial indicators from different sentences shows similar semantics. If so, we can conclude that this context representation of a spatial indicator does represent its true spatial function. To perform this check, we collect the context vectors of the spatial indicators from all training sentences and cluster them using average linkage clustering with an empirically decided distance threshold of 0.7. A manual observation of the clusters shows that the same indicator from different sentences lies mostly in a single cluster and different spatial indicators lie in their respective clusters, thus validating the proposed understanding.

5 Conclusion

In this paper, we attempt to solve the problem of identification of spatial roles and static spatial relations in text. We show that context of words learned from a BiLSTM trained for a sequence labelling task can help in the identification process. We show that our two-step approach of generating context vectors and relation identification based on the learned context vectors, outperforms the state-of-the-art results on tasks of SemEval 2012 and SemEval 2013.

References

1. Bastianelli, E., Croce, D., Basili, R., Nardi, D.: Unitor-hmm-tk: Structured kernel-based learning for spatial role labeling. In: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). vol. 2, pp. 573–579 (2013)
2. Chollet, F.: keras. GitHub repository. https://github.com/fchollet/keras. Accessed on 26-October-2018
3. Grubinger, M., Clough, P., Müller, H., Deselaers, T.: The iapr tc-12 benchmark: A new evaluation resource for visual information systems. In: Int. Workshop OntoImage. vol. 5 (2006)
4. Kolomiyets, O., Kordjamshidi, P., Moens, M.F., Bethard, S.: Semeval-2013 task 3: Spatial role labeling. In: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). vol. 2, pp. 255–262 (2013)
5. Kordjamshidi, P., Bethard, S., Moens, M.F.: Semeval-2012 task 3: Spatial role labeling. In: Proceedings of 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. pp. 365–373. Association for Computational Linguistics (2012)
6. Kordjamshidi, P., Moens, M.F.: Global machine learning for spatial ontology population. Web Semantics: Science, Services and Agents on the World Wide Web **30**, 3–21 (2015)
7. Ludwig, O., Liu, X., Kordjamshidi, P., Moens, M.F.: Deep embedding for spatial role labeling. arXiv preprint arXiv:1603.08474 (2016)
8. Mazalov, A., Martins, B., Matos, D.: Spatial role labeling with convolutional neural networks. In: Proceedings of the 9th Workshop on Geographic Information Retrieval. p. 12. ACM (2015)
9. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 1532–1543 (2014)
10. Pustejovsky, J., Kordjamshidi, P., Moens, M.F., Levine, A., Dworman, S., Yocum, Z.: Semeval-2015 task 8: Spaceeval. In: Proceedings of the 9th International Workshop on Semantic Evaluation (semeval 2015). pp. 884–894 (2015)
11. Pustejovsky, J., Moszkowicz, J.L., Verhagen, M.: Iso-space: The annotation of spatial information in language. In: Proceedings of the Sixth Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation. vol. 6, pp. 1–9 (2011)
12. Roberts, K., Harabagiu, S.M.: Utd-sprl: a joint approach to spatial role labeling. In: Proceedings of 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. pp. 419–424. Association for Computational Linguistics (2012)