



Spatial Role Labeling based on Improved Pre-trained Word Embeddings and Transfer Learning

Alaeddine Moussa, Sébastien Fournier, Khaoula Mahmoudi, Bernard Espinasse,
Sami Faiz

► To cite this version:

Alaeddine Moussa, Sébastien Fournier, Khaoula Mahmoudi, Bernard Espinasse, Sami Faiz. Spatial Role Labeling based on Improved Pre-trained Word Embeddings and Transfer Learning. *Procedia Computer Science*, 2021, 192, pp.1218-1226. <10.1016/j.procs.2021.08.125>. <hal-03594477>

HAL Id: hal-03594477

<https://hal.science/hal-03594477v1>

Submitted on 16 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License



25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Spatial Role Labeling based on Improved Pre-trained Word Embeddings and Transfer Learning

Alaeddine Moussa^{a,b,c,*}, Sébastien Fournier^a, Khaoula Mahmoudi^c, Bernard Espinasse^a, Sami Faiz^c

^aAix Marseille Université, CNRS, LIS UMR 7020

^bUniv. Manouba, ENSI, Campus Universitaire Manouba

^cUniversité de Tunis El Manar, ENIT, Laboratoire de Télédétection et Systèmes d'Information à Référence Spatiale (UR-TSIRS-ENIT)

Abstract

In several real-world applications, extracting spatial semantics from text is critical. Spatial Role Labeling (SpRL) introduces a language-independent annotation scheme used in these applications, particularly for reasoning purposes. This paper proposes, first of all, a transfer learning method with a word embeddings-based approach for SpRL. Then, we enhance the word vectors with POS tags and CNN-based character-level representations. Finally, we propose a Residual BiLSTM CRF deep learning model to identify the spatial roles. The experimental results on two datasets: SemEval-2012 and SemEval-2013 Task 3, show that the proposed model outperforms other machine learning approaches.

© 2021 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of the KES International.

Keywords: Spatial Role Labeling; Word Embedding; Transfer Learning; Deep Learning; Natural language processing

1. Introduction

Spatial Role Labeling (SpRL) [11] introduces a language-independent annotation scheme to identify spatial scene objects and relationships within a text, including reasoning purposes. Spatial Role Labeling is used in many application areas such as robotics [16], maritime navigation [7], traffic management [1], in query response systems [3]. It involves processing sentences in a text and identifying the objects in spatial scenes and their relationships.

Linguistic constructs may communicate complex relationship structures and spatial relationships between them and movement patterns across space relative to a reference point. A spatial relationship between two objects is typically expressed in natural language through a preposition (e.g. in, on, at, ...) or prepositional expression (on top of, inside of, ...).

* Corresponding author.

E-mail address: alaeddine.moussa@etu.univ-amu.fr



Fig. 1. Example sentences annotated with SpRL-2013 annotation.

For instance, we show in Figure 1 a sentence with a set of words directed by the token "cup" refers to a trajectory (TR) object, the group of words directed by the token "table" refers to the role of a landmark (LM), and these are linked by the spatial indicator (SI) "on." The spatial indicator (often a preposition) establishes the type of spatial relationship. Thus, the task of SpRL is to process the sentences of a text automatically and identify the spatial scene objects and the relations between them.

A trajector (TR) describes a central object of a spatial scene. It can be static or dynamic; including persons, objects, or events. The landmark (LM) is a spatial role label that denotes the location or the motion of the TR. It indicates a secondary object of a spatial scene, to which a possible spatial relation can figure out. A spatial Indicator (SI) is a spatial role label allocated to a word or a phrase to flags a spatial relation between objects (TR and LM).

Most of the existing SpRL approaches imply Word Embedding (WE). WE is one of the most useful deep learning methods used for constructing vector representations of words. Although very effective, these methods have certain limits and need to be improved. WE need large corpora for training and presenting a suitable vector for each word.

In this research, we propose to improve the accuracy of pre-trained WE. Given the small size of the SpRL corpus, we use transfer learning to transport the knowledge from a large corpus to a small one. Then, by combining the word vector with a bag of features. The results show that proposed method increases the accuracy of pre-trained word embeddings vectors for SpRL. The main contributions of our work are the following:

- 1) We present a transfer learning approach for SpRL.
- 2) We combine the WE vector with bag features, including Part of Speech and character-level representation.
- 3) We use the well-known state-of-art deep learning architecture BiLSTM-CRF with Residual connection for SpRL.

The remainder of this paper is organized as follows : Section 2 describes the related works and literature review. Section 3 presents our proposed method and describes the proposed deep learning model. Section 4 reports the designed experimental setup. Results are discussed in Section 5. Finally, Section 6 summarizes the major findings.

2. Related work

Several research studies have focused on SpRL. SemEval is a series of ongoing evaluations of computational semantic analysis systems to explore the nature of meaning in language. Although meaning is intuitive to humans, the transfer of these intuitions to computer analysis has proven complicated.

The SemEval campaigns in automatic information extraction of the years 2012, 2013, and 2015 involved challenges to be taken up in SpRL on a specific proposed corpus of reference texts. Thus, the SpRL task at SemEval 2012 [9] focused on the roles of TRs, LMs, SIs, and the links between these roles form spatial relations. The formal semantics of the relations were considered at a granular level, composed of three types of relations: directional, regional (topological), and distal (the part furthest from a reference center). The spatial roles are assigned both to phrases and their headwords, but only the headwords are evaluated for this task.

In the SemEval 2013 SpRL task, [8] introduced new roles to include motion indicators, paths, directions, and distances to capture the fine-grained spatial semantics of static spatial relationships and also to take into account dynamic spatial relationships. In this task, the entire span of spatial roles is evaluated, not only headwords.

The SpaceEval task, introduced in SemEval-2015, adopted a more advanced annotation specification with respect to ISOSpace [18]. SpaceEval, first, enriches the semantics' granularity in both static and dynamic spatial configurations, and secondly, by extending the variety of annotated data and the domains considered. Indeed, the concept of place is distinguished from the concept of a spatial entity.

The KUL-SKIP-CHAIN-CRF [11] system adopted a pipeline approach. In the beginning, spatial indicators are found and labeled for each sentence. Then, given a spatial indicator, the second task consists of classifying parts of an input sentence with LM or TR labels. To identify the SI, authors use an external corpus The Preposition Project (TPP) labeled to learn the sense of preposition through a set of linguistic features (lemma, POS,...). As a result, the binary classification of a preposition's spatial is made. In the same way, TR and LM are identified to leverage many linguistically motivated features. Finally, SIs, TRs, and LMs are combined into spatial relation triplets.

The UTDSpRL [19] system used a join approach to tackle the SpRL task. Authors expose that the pipeline approach cannot perform well if a spatial roles arguments are considered in isolation. They propose an alternative system that jointly decides whether a given candidate triple expresses a spatial relation or not. The approach used a recall heuristics to find spatial relation candidate triple. A hand-crafted dictionary was used to detect SPATIAL_INDICATOR candidates, allowing the decrease of negative relation candidates in comparison to the pipeline method, which considers every phrase as a spatial role. Then, noun phrase heads were treated as TRAJECTOR and LAND-MARK candidates. Finally, a Support Vector Machine (SVM), was trained with some manually engineered features chosen by an automatic feature selector, to classify the relation candidate.

The UNITOR-HMM-TK [2] system uses a sequence-based classifier. First, classifies spatial and motion indicators, then built on these outcomes to identify spatial roles through a set of lexical and grammatical features. The generated candidate spatial relations are verified by a Support Vector Machine (SVM). This approach's significant contribution is to adopt smooth grammatical features instead of a full syntax of the sentence.

SpRL-CNN system [17] rely on an adapted version of the nlpnet¹ system. First, a spatial indicator is identified by a multilayer perceptron (MLP) which it converts the tokens (spatial indicator) into feature vectors. Then, these vectors are fed to Conventional Neural Network (CNN) to classify the argument and spatial relation triples. Additionally, pre-trained Glove word embeddings were used to reduce the impact of words not seen in the training data.

The VIEW system [13] consists of producing word embedding vectors from multimodal. The goal is to be able to encode the visual information of images within the "Feature vector" produced by [10] by concatenating them with the "embedding" of VIEW.

Recently, [5] extended SpRL to encode data in radiology context. They presented a dataset with 2000 chest X-ray reports. The annotation involves a radiographic conclusion and its associated anatomical position.

3. Proposed method

In what follows we present our step-wise approach based on the combination of natural language processing techniques, Word2Vec methods, and transfer learning techniques. Figure 2 sketches our approach major steps.

3.1. Transfer learning

Transfer learning involves all methods that utilize any supplementary resources to enhance model learning for the target task. Previous SpRL systems did not focus on the importance of text vectoring, mainly because the size of the SpRL corpus is small. Moreover, the quality of the word vectors is proportionally related to the amount of data. The larger the corpus, the better the vectors are. Our idea is to generate word vectors from a database more prominent than the SpRL one. We have chosen the MS COCO (Microsoft Common Objects in Context) database². This choice is motivated by the strong resemblance between the contents of the two databases. For example, a typical image caption from MS COCO would be "A large bus sitting next to a very tall building." Indeed, The MS COCO dataset is large-scale object detection, segmentation, key-point detection, and captioning dataset. The dataset consists of 328K images and five reference captions for every image. We use those image captions to generate word vectors.

¹ <http://nilc.icmc.usp.br/nlpnet/>

² <https://cocodataset.org>

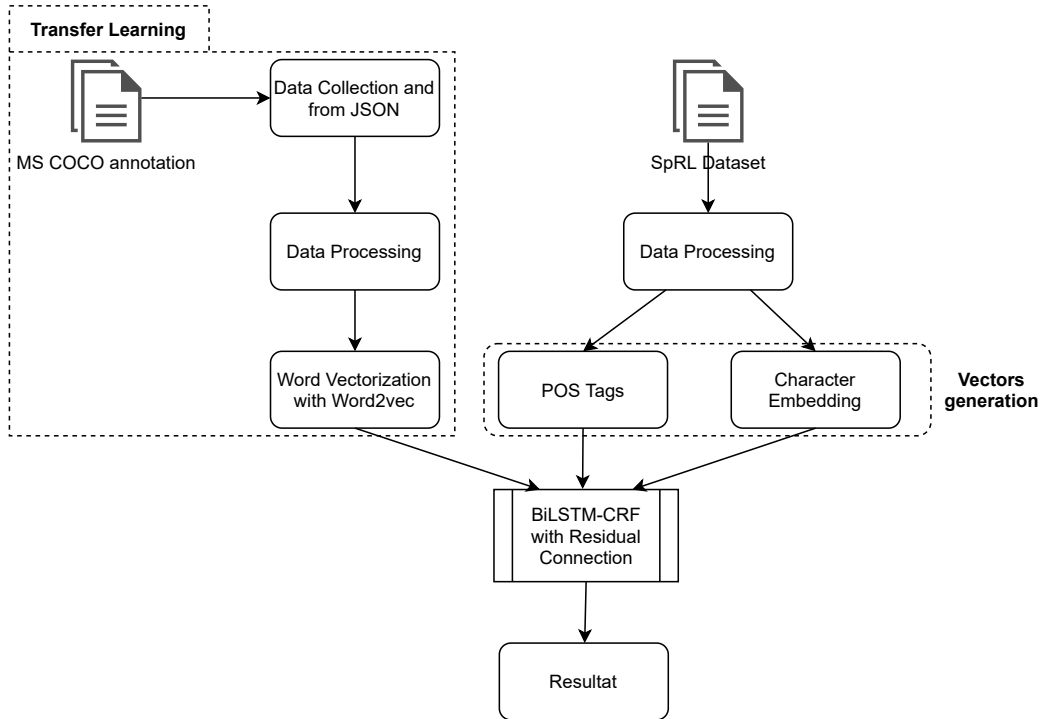


Fig. 2. An illustration of Residual BiLSTM-CRF for Spatial Role Labeling.

3.2. Part-of-speech to vector

Part-of-speech (POS) tagging is an essential and fundamental step in Natural Language Processing, indicating each word of a text the proper POS tag. The POS gives a large amount of information about a word and its neighbors, syntactic groups of words (nouns, verbs, adjectives, adverbs), and similarities and differences. We converted each generated POS tag to a vector representation. As a result, Word2Vec vectors will have syntactic information of words.

3.3. Character to vector

Character-level contains much information about the meaning of the word. It proves that it can memorize the arbitrary aspects of word spelling. Using this information can significantly improve the quality of the model [4]. To encode the structure of words, we will use a Convolution Neural Network (CNN). We will encode each character into a vector using one-hot encoding and feed them into a CNN.

3.4. Residual-BiLSTM-CRF model

Our neural network architecture is based on the BiLSTM-CRF model [14] modified to incorporate two BiLSTM layers and a residual connection between them. When processing sequence labeling tasks, each labeling decision for each word is produced independently of the others, but in many cases neighboring decisions leverage each other. The Bi-directional LSTM (BiLSTM) encodes each sequence forwards and backward to capture past and future information. The LSTM architecture is described in mathematical terms for a unit at time j as:

$$c_j = f \odot c_{j-1} + i \odot z \quad (1)$$

$$h_j = o \odot \tanh(c_j) \quad (2)$$

$$i = \sigma(x_j W_{xi} + h_{j-1} U_{hi}) \quad (3)$$

$$f = \sigma(x_j W_{xf} + h_{j-1} U_{hf}) \quad (4)$$

$$o = \sigma(x_j W_{xo} + h_{j-1} U_{ho}) \quad (5)$$

$$z = \tanh(x_j W_{xz} + h_{j-1} U_{hz}) \quad (6)$$

where c_j denotes the memory cell and h_j is the hidden state. σ denotes the element-wise sigmoid function, and \odot is the element-wise product. i , f , and o are the three gates. W_{xi} , W_{xf} , W_{xo} denote the weight matrices of different gates for input x_j , and U_{hi} , U_{hf} , U_{ho} are the weight matrices for hidden state h_j . The values of the gate are determined on the basis of linear combinations of the current input x_j and the previous state h_{j-1} using a sigmoid function. The updated candidate z is estimated as a linear combination of x_j and h_j via a tanh function.

Residual LSTM states that separating a spatial-domain shortcut path with a temporal-domain cell update gives better flexibility to deal with vanishing gradients. The following equation computes the output of each sub-layer:

$$z = \text{LayerNorm}(x + \text{Sublayer}(x)) \quad (7)$$

where *LayerNorm* is an affine function that stabilizes the activation of a deep neural network.

Given an input sequence $x = (x_1, \dots, x_n)$ and a sequence of SpRL labels $y = (y_1, \dots, y_n)$ for x . The probabilistic model for overall possible tag sequences will be defined as:

$$p(s|x; w) = \frac{e^{(w \cdot \Phi(x, y))}}{\sum_{x'} e^{(w \cdot \Phi(x', y))}} \quad (8)$$

where s extends overall possible output sequence and w is the parameter vector. Indeed, the expression $w \cdot \Phi(x, y) = \text{score}_{crf}(x, y)$ represents the scoring function that indicates how y fits x . In the CRF layer, we aim to replace the linear scoring function with a non-linear neural network. We define the score by:

$$\text{score}_{BiLSTM-crf}(x, y) = \sum_{i=0}^n W_{y_{i-1}, y_i} \cdot BiLSTM(x)_i + b_{y_{i-1}, y_i} \quad (9)$$

With the score function constructed, we can efficiently use past and future tags to predict the current tag and optimize the conditional probability $p(y|x; w, b)$ and propagating back through the network.

The SpRL task is divided mainly into two sections: spatial role identification and spatial relation identification tasks. For the relation extraction task, we aim to figure out which spatial roles discovered in the previous classification phase can be composed as valid spatial relations. All possible spatial roles are first generated by heuristics and then combinatorially combined to acquire candidate relationships. The network was feed with three input representation vectors v , each referring to the TR, SI, and LM.

4. Experiments

We lead experiments to evaluate the performance of the proposed approach for SpRL. In this section, we report datasets used for testing, followed by the experimental setup.

4.1. Datasets and preprocessing

We use the SemEval-2012 and SemEval-2013 shared tasks. We have chosen these two corpora because they are the reference corpora for SpRL and most systems have been tested in.

4.1.1. SemEval-2012 Data

The first corpus is a subset of the IAPR TC-12 image Benchmark [6]. It contains 613 text files that include 1213 sentences in total. The original corpus was available without copyright restrictions³. Tourists took 20,000 images with textual descriptions in different languages. The texts describe objects in a scene together with their absolute and relative positions in the image. The data is released in XML format.

4.1.2. SemEval-2013 Data

The dataset for SemEval2013 includes two distinctive corpora. The first one is the SemEval2012 Data presented in the previous sub-section but with some modifications. The Data was transformed into a span based annotation, and some annotation error was corrected. The second corpus originates from the Confluence Project⁴. This project tries to outline all possible latitude-longitude intersections on earth, and people who visit these intersections provide written narratives of the visit.

In this workshop, systems were evaluated mainly into tow tasks : Individual role identification and spatial relation extraction. For SemEval2013 task 3, all reported results followed the relaxed evaluation criteria as motioned by the organizers.

4.1.3. Preprocessing

The original XML encoding for the previous datasets was converted into CoNLL-U format with Stanford CoreNLP toolkit [15] to get POS tag. We added a new column that holds the tag for each word in the IOBES labeling scheme.

Concerning the MS coco dataset, the annotations of the images were in the form of a JSON file. We have created a Java program that parses the file and creates a raw textual corpus with the annotations only.

4.2. Experimental Setup

For Word Embeddings, we employ the word2vecf toolkit⁵ [12] to generate word vectors from MS COCO.

We implement the neural network using the Keras library. We choose 1D Convolutional Neural Networks (CNN) to construct character-level embeddings followed by a 1d maxpool operation. Hyper-parameters of the CNN was set after tuning to an embedding size of 30, a window size of 3, and 30 filters. We have tested LSTM-based character-level embeddings, which achieved similar performance but, the CNN embeddings reduced training complexity. The size of the POS tag came empirically with 30 dimensions each. Besides, The CRF layer was implemented with the Keras-contrib package.

We optimize our model by mini-batch stochastic gradient descent (SGD) with a learning rate of 0.01 and batch size 10. Also, we include the L2 regularization with parameter 1e-8.

³ <http://www.cs.tulane.edu/~pkordjam/SpRL.htm#data>

⁴ <http://confluence.org>

⁵ <https://bitbucket.org/yoavgo/word2vecf>

Table 1. F1 scores comparison on the SemEval 2012 dataset with TR=Trajector, LM=Landmark, SI=Spatial Indicator, and relation=Spatial relation

System	TR	LM	SI	Relation
W2V _{SpRL} -BiLSTM-CRF	0.633	0.685	0.887	0.601
W2V _{MS COCO} -BiLSTM-CRF	0.652	0.731	0.892	0.610
W2V _{MS COCO} -BiLSTM-CRF+POS+char	0.700	0.750	0.902	0.622
W2V _{MS COCO} -BiLSTM-CRF _{Residual} +POS+char	0.710	0.783	0.905	0.638
----- SpRL systems -----				
EtoE-IBT-CLCP [10]	0.673	0.797	0.869	0.617
KUL-SKIP-CHAIN-CRF [11]	0.646	0.756	0.900	0.500
UTDSpRL [19]	0.707	0.772	0.823	0.573

Table 2. F1 scores comparison on the SemEval 2013 dataset with TR=Trajector, LM=Landmark, SI=Spatial Indicator, and relation=Spatial relation

System	TR	LM	SI	Relation
W2V _{SpRL} -BiLSTM-CRF	0.613	0.652	0.831	0.427
W2V _{MS COCO} -BiLSTM-CRF	0.628	0.661	0.875	0.438
W2V _{MS COCO} -BiLSTM-CRF+POS+char	0.702	0.753	0.899	0.580
W2V _{MS COCO} -BiLSTM-CRF _{Residual} +POS+char	0.735	0.791	0.920	0.604
----- SpRL systems -----				
SpRL-CNN [17]	NA	NA	NA	0.460
UNITOR-HMM-TK [2]	0.682	0.785	0.926	0.458
VIEW [13]	0.732	0.678	0.749	0.235

5. Results and discussion

In this section, the experimental results are shown and discussed. First of all we present an ablation analysis of results, then an error analysis of experimental results. We present our main results for SpRL individual spatial role identification in Table 1 and Table 2.

Systems were evaluated against the original gold annotations. A spatial role (TR, LM or SI) is considered correct if it has a minimal overlap of one character with a gold annotation and matches the role type of the gold annotation. For spatial relation tuple, we considered the relation correct if it is of the same length then the gold annotation and if each single role in the system tuple matches each role in the gold tuple.

The evaluation metrics of precision, recall and F1 measure are used, and are defined as:

$$precision = \frac{TP}{TP + FP} \quad (10)$$

$$recall = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = \frac{2 * recall * precision}{recall + precision} \quad (12)$$

Where TP is the number of predicted tags that precisely match the ground truth, FP is the number of predicted tags that do not match the ground truth, and FN is the number of ground truth tags that do not match the predicted components.

5.1. Ablation study

In order to observe the effects of transfer learning on our model, we conducted an ablation analysis. We compared the use of WE trained locally on the SpRL-2012 corpus against the WE trained on MS COCO. Table 1 and Table 2 shows that WE trained on MS COCO led to better results.

Based on those results, we can deduce the influence of supplementing a component to the final architecture. We observe that adding POS tags and character representation improves the performance significantly compared to only

using the words. Also, we notice increases when we add residual connections between each layer. The residual connections let the model decide whether it needs to go deeper or skip the stacked layers at its convenience.

5.2. Error analysis

To understand the nature of the errors, we manually inspected them. One common error is assigning none tag to a token that plays spatial roles. For example, in the following sentence,

Well exposed front view of the cathedral in Asuncion.

”cathedral” is wrongly classified as none.

Otherwise, we may find ourselves with errors in which the classifier assigns a space role to a word that is not one. In this example,

A flat landscape with green trees and houses in the background.

the word ”trees” is wrongly classified as Trajectory.

6. Conclusion

In this paper, we proposed a new method to improve pre-trained word embeddings for SpRL. First of all, we use a transfer learning approach to generate and fine-tune the word vectors. Then, we enhance the word vectors using both POS tags and CNN-Based character representation. In the next step, we combine the LSTM network with residual connections to help the identification of the critical features from the embedded vector. Finally, we add a CRF layer to exploit the neighbor tags information while predicting current tags at the last stage.

Experimentation results are conducted on SemEval-2012 and SemEval-2013 Task 3 datasets and show that our proposed method increases the accuracy of Spatial information classification tasks in all datasets.

For future work, we ought to improve the performance of our work by introducing syntactic dependencies in Word Embedding and mixing them with pre-trained ones.

References

- [1] Aydemir, F.B., Dalpiaz, F., 2017. Towards aligning multi-concern models via nlp, in: 2017 IEEE 25th International Requirements Engineering Conference Workshops (REW), IEEE. pp. 46–50.
- [2] Bastianelli, E., Croce, D., Nardi, D., Basili, R., 2013. UNITOR-HMM-TK : Structured Kernel-based Learning for Spatial Role Labeling, in: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013), pp. 573–579. URL: https://www.cs.york.ac.uk/semeval-2013/accepted/52f_}Paper.pdf.
- [3] Bereta, K., Xiao, G., Koubarakis, M., 2019. Ontop-spatial: Ontop of geospatial databases. Journal of Web Semantics 58, 100514.
- [4] Chen, X., Xu, L., Liu, Z., Sun, M., Luan, H., 2015. Joint Learning of Character and Word Embeddings , 1236–1242.
- [5] Datta, S., Roberts, K., 2020. A dataset of chest X-ray reports annotated with Spatial Role Labeling annotations. Data in Brief 32, 103473. URL: <https://doi.org/10.17632/yhb26hfz8n.1>, <https://github.com/krobertslab/datasets/tree/master/rad-sprl>, doi:10.1016/j.dib.2020.106056.
- [6] Grübinger, M., Clough, P., Müller, H., Deselaers, T., 2006. The IAPR TC-12 Benchmark: A New Evaluation Resource for Visual Information Systems, in: LREC Workshop OntoImage Language Resources for Content-Based Image Retrieval, pp. 13–23. URL: <http://www-i6.informatik.rwth-aachen.de/publications/download/34/TheIAPRBenchmark:ANewEvaluationResourceforVisualInformationSystems.pdf>.
- [7] Hadzagic, M., Isabelle, M., Kashyap, N., 2020. Hard and soft data fusion for maritime traffic monitoring using the integrated ornstein-uhlenbeck process, in: 2020 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), IEEE. pp. 98–105.
- [8] Kolomiyets, O., Kordjamshidi, P., Moens, M.F., Bethard, S., 2013. SemEval-2013 Task 3: Spatial Role Labeling. Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013) 2, 255–262. URL: <http://www.aclweb.org/anthology/S13-2044>.
- [9] Kordjamshidi, P., Bethard, S., Moens, M.F., 2012. SemEval-2012 Task 3: Spatial Role Labeling. Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval-2012) , 365–373 URL: <https://dl.acm.org/citation.cfm?id=2387694>.
- [10] Kordjamshidi, P., Moens, M.F., 2015. Global machine learning for spatial ontology population. Journal of Web Semantics 30, 3–21. URL: <http://dx.doi.org/10.1016/j.websem.2014.06.001>, doi:10.1016/j.websem.2014.06.001.

- [11] Kordjamshidi, P., Van Otterlo, M., Moens, M.F., 2011. Spatial role labeling. *ACM Transactions on Speech and Language Processing* 8, 1–36. URL: <http://dl.acm.org/citation.cfm?doid=2050104.2050105>, doi:10.1145/2050104.2050105.
- [12] Levy, O., Goldberg, Y., 2014. Dependency-Based Word Embeddings. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* , 302–308 URL: <http://aclweb.org/anthology/P14-2050>, doi:10.3115/v1/P14-2050.
- [13] Ludwig, O., Liu, X., Kordjamshidi, P., Moens, M.F., 2016. Deep Embedding for Spatial Role Labeling , 1–19 URL: <http://arxiv.org/abs/1603.08474>, doi:10.13140/RG.2.1.5080.4884, [arXiv:1603.08474](https://arxiv.org/abs/1603.08474).
- [14] Ma, X., Hovy, E., 2016. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF, in: 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers, pp. 1064–1074. doi:10.18653/v1/p16-1101, [arXiv:1603.01354](https://arxiv.org/abs/1603.01354).
- [15] Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, D., 2014. The Stanford CoreNLP Natural Language Processing Toolkit, in: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Association for Computational Linguistics, Stroudsburg, PA, USA. pp. 55–60. URL: <http://aclweb.org/anthology/P14-5010>, doi:10.3115/v1/P14-5010.
- [16] Matuszek, C., 2018. Grounded language learning: Where robotics and nlp meet (invited talk), in: *Proceedings of the International Joint Conference on Artificial Intelligence*.
- [17] Mazalov, A., 2016. Understanding Spatial Semantics in Natural Language URL: <https://fenix.tecnico.ulisboa.pt/downloadFile/1689244997257025/Extended{ }Abstract{ }Mazalov{ }80462.pdf>.
- [18] Pustejovsky, J., Kordjamshidi, P., Moens, M.f., Levine, A., Dworman, S., Yocum, Z., Leuven, K.U., 2015. SemEval-2015 Task 8: SpaceEval , 884–894.
- [19] Roberts, K., Harabagiu, S.M., 2012. UTD-SpRL: A Joint Approach to Spatial Role Labeling , 419–424 URL: <http://www.aclweb.org/anthology/S12-1056>.