Saarland University
Language Science and Technology
Master's Thesis
Prof. Dr. Dietrich Klakow
Prof. Dr. Iryna Gurevych

Relation Extraction Using Liberalism love and beloved

December 1, 2013

Ehsan Khoddammohammadi

Acknowledgements

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Declaration

I hereby confirm that the thesis presented here is my own work, with all assistance acknowledged.

Saarbrücken, 4th December 2013

Ehsan Khoddammohammadi

Abstract

Keywords: Blah, Blah

Contents

1 Entity Linking Among Lexical Resources

Each KB or lexical resource contains limited amount of information and different structured resources have different perspective and partially encode information about an entity and its relations to the other entities of the resource. Therefor, if one is interested about learning different features from a lexical resource or a KB, a natural next step is to jointly learn features from multiple resources. In previous chapter we saw how to learn features from a KB and a corpus with application to relation discovery and extraction. In this chapter we focus on inducing features for entities of multiple structured KBs. For relation extraction, the most relevant available resource is Freebase and it suffices together with corpus to learn features of named entities (NE). In this chapter we focus on learning features for other type of words e.g. nouns, verbs and ... . Learning features for other words in a sentence rather than just NEs might help improving the performance of a relation discovery system specially for verbs (relations) when some semantic aspects of them like selectional preference has been shown to be an important aspect. Another application is learning and linking entities of lexical resources in two different languages which helps to transfer information from one language domain to the another one. For these reasons, first we describe a pipeline to learn word features from multiple resources and as proof of concept I implement this idea with inducing bi-lingual word features and show its performance on semantic similarity task. The actual application of this idea and learned word features is postponed to future work which I describe potential usage of them in relation discovery.

1.1 Task Description

In this part I describe the methodology we followed to encode available information in two different lexical resources, WordNet and GermaNet, that makes it possible to link entities of the two different resources and learn bi-lingual embeddings of word senses in

German and English. The main idea is to relate two senses from two different resources using cross-lingual sense alignments. This is an additional information which can play a role of bridge between two different tasks, learning German embeddings and English embeddings, and can help to transfer knowledge from one to the another. Using this new feature we make our WordNet-GermaNet dataset which contains three types of relations (1) WordNet relations (2) GermaNet relations (3) Cross-lingual sense alignments between WordNet and GermaNet

First two types of relations are directly extracted from WordNet and GermaNet and for the cross-lingual relations we used Interlingual Index mappings between WordNet and GermaNet.
Example of relations:

WN-sense-A	WN-rel-1	WN-sense-B
GN-sense-C	GN-rel-2	GN-sense-D
WN-sense-A	ILI-rel-1-2	GN-sense-D

Left and right entities are WordNet and GermaNet senses and relations are current semantical relations in each of lexicons such as: meronymy, holonymy and ....

1.2 Experimental Setup

We have created four different dataset, each divided to train, test and validation separated subsets. Our four datasets are:

1. Only WordNet triples (WN)
2. Only GermaNet triples (GN)
3. WordNet-GermaNet triples with one-direction cross-lingual alignments (WN-GN)
4. WordNet-GermaNet with double-direction cross-lingual alignments (WN-GN DD)

Dataset 3 includes both relations extracted from WordNet and GermaNet and also the mapping between senses. Dataset 4 is same as dataset 3 but since the models we will use are assuming all the relations are assymetric we will try to encode the symmetry of cross-lingual alignments by reversing each of them and include the reverse in the dataset. Datasets 3 and 4 contains two different variants: the first variants contains only WordNet relations (test on WN) in the held-out test dataset and the second variant contains only GermaNet triples (test on GN). In this way we can observe the direction of possibly transferring information from English to German or vice versa.

For reducing the sparsity of data and boosting the learning runtime we filtered out all the entities that appeared less than 3 times in our datasets.

(version of wordnet, germanet, ILI and role of uby should be described here)

1.3 Evaluation

To show the effectiveness of joint learning of features from multiple knowledge bases we suggest two experiment setups. In the first schema we follow Bordes et al. ranking task. The goal of this task is to show how well the information in knowledge bases can be preserved by the learned features. On the other hand, the second setup is investigating on this question that if the learned word embeddings from multiple resources are able to improve the performance of monolingual embeddings in a standard NLP task, here word-pair similarity or not. In this setup we will look to contribution of the learned features in predicting similarity of words.

1.3.1 Evaluation Using Reconstruction

Bordes et al. (Bordes2011) proposed a ranking task that for each triple (e_i, r_k, e_j) in the data set, all the entities will be ranked as a candidate for being right entity of the triple given the relation and the left entity. Depends on which one of the models is used, SE or SME-Bil, all the entities will be sorted based on their score regarding Equation (**??**) or Equation **??** previously introduced in section **??**. By keeping the statistics of difference between the predicted rank of e_j and its true rank and also repeat the same process for left entities, we will be able to report the mean and median predicted rank of entities per relation and in total. Bordes et al. proposed to schema for calculating the average rank, micro averaging which emphasis on more frequent relations by weighted averaging with frequency of relations as weights and macro averaging which consider all the relations equally , either frequent or infrequent ones. The third statistic that we report following their work, r@100, is the ratio of number of times that an entity is correctly among top 100 entities ranked and predicted for a triple to the number of occurrences of this entity in the dataset. We applied SE and SME-Bil models on our created datasets and the ranking performance on each of them is presented in Table 1.1.

1.3.2 Evaluation on Semantic Similarity

We are interested to further analyze the effectiveness of learned embeddings to capture semantic features of words, therefore we compare the mono-lingual and bi-lingual embeddings against human judgments and also other embeddings learned from corpus. The other embeddings which we used for our comparison are (Turian et al., HLBL and Klementiev et al.). To measure the similarity between any given wordpair (w_1, w_2) we find all vectors associated to different senses of the given words in our embedding dictionary and pick the pair of embeddings that maximize cosine similarity between two words. We can motivate this by saying that for each word pair any of words works as a context for disambiguating the sense of the other word.

Four datasets of word-pair similarity are used to compare correlation of predicted similairty of pair of words against human judgments. RG-65 [rubensteinGoodenough], Yang and Powers verb similarity dataset[yangPowers], MC-30[millerCharles] and WS-353 [finkelstein] are English datasets that we used for this task. RG-65 and its subset, MC-30 are providing human scored datasets for measuring synonymy among wordpairs (nouns),. WS-353 has broader notaion of semantic similarity and include word pairs for measuirng semantic relatedness too. Yang and Powers have provided a dataset for measuring semantic similarity between verbs.

Both Pearson and Spearman correlation of predicted and gold similarities are calculated and is reported in table 1.2 for English. We can see that bi-lingual embeddings learned by SME-Bil model outperformed all other embeddings both in Pearson and Spearman correlation in all four datasets. Another observation is that SME-Bil models performes better than SE models in most cases. Among SME-Bil models, bi-lingual embeddings are always more correlated to human hudgments than mono-lingual embeddings. All models perform poorly on WS-353, bi-lingual SME-Bil model still performs better than the rest. We believe this could due to including notion of relatedness in this dataset. Poor performance of WordNet based models on semantic relatedness task is previously known and discussed in [**?**][Szulmanski2013]. Words can be related by different type of relations while in WordNet, a few number of relations are encoded e.g. synonymy which is the subject of our evaluation task.

For German, we use translated versions of RG-65, MC-30 and WS-353 which were judged by German native speakers [**?**] [gurevych et al], additionally a dataset from [**?**] [Gurevych et al.] is used. The structured embeddings are being compared to the only available German word embeddings from [1][Klem et al.]. We can observe that bi-lingual embeddings always outperform mono-lingual embeddings except for RG-DE-65 dataset which both mono-lingual and bi-lingual **SE** models have equal Spearman correlation to the gold data but mono-lingual model performs better in Pearson correlation. The good results of German evaluation is another indicator for proofing the idea of learning word embeddings from multiple resources.

1.4 Conclusion

Table 1.1: Intrinsic Evaluation (Ranking Score Performance)

Dataset	#relations	#entities		Micro	Macro
GN SE	16	64025	mean	1003.59	3739.85
			median	5.0	2213.37
			global	84.23	72.49
GN SME-Bil	16	64025	mean	407.90	308.01
			median	10.0	54.18
			global	81.18	69.85
WN SE	23	148976	mean	148.72	623.10
			median	5.0	4.69
			global	92.10	89.86
WN SME-Bil	23	148976	mean	128.82	511.21
			median	10.0	26.63
			global	84.14	75.57
WN-GN SE (WN held out)	32	213002	mean	293.16	1356.30
			median	5.0	5.10
			global	91.19	88.95
WN-GN SME-Bil(WN held out)	32	213002	mean	124.85	331.82
			median	11.0	33.86
			global	82.91	73.55
WN-GN SE (GN held out)	32	213002	mean	3031.44	15470.56
			median	7.0	10080.5
			global	80.87	70.313
WN-GN SME-Bil(GN held out)	32	213002	mean	984.79	1021.37
			median	40.0	428.90
			global	64.16	55.98
WordNet-GermaNet-DD SME-Bil (WN held out)	32	213002	mean	166.18	466.91
			median	18.0	55.41
			global	77.07	65.082
WordNet-GermaNet-DD SME-Bil (GN held out)	32	213002	mean	932.49	719.47
			median	56.0	175.56
			global	59.22	50.84

Table 1.2: Word-pair Similarity Performance for English

Dataset		WN-SE	WN-GN-SE	WN-SME-Bil	WN-GN-SME-Bil	WN-GN-SME-Bil-DD	HLBL	Turian et al.	Klementiev et al.
RG-65	P	0.682	0.666	0.703	0.833	0.725	-0.115	0.233	-0.380
	S	0.769	0.741	0.741	0.811	0.825	-.083	0.118	-0.398
MC-30	P	0.611	0.644	0.601	0.740	0.599	-0.363	0.150	-0.768
	S	0.720	0.648	0.756	0.846	0.954	-.450	-0.198	-0.522
WS-353	P	0.181	0.206	0.239	0.246	0.238	0.233	0.236	0.029
	S	0.093	0.146	0.185	0.224	0.201	0.197	0.210	0.040
YangPowers-130	P	0.482	0.637	0.584	0.627	0.610	-0.130	-0.076	0.154
	S	0.401	0.472	0.406	0.553	0.533	-0.186	-0.116	0.113

Table 1.3: Word-pair Similarity Performance for German

Dataset		GN-SE	WN-GN-SE	GN-SME-Bil	WN-GN-SME-Bil	WN-GN-SME-Bil-DD	Klementiev et al.
wortpaare222	P	-0.010	0.156	0.073	0.130	0.196	0.107
	S	-0.125	0.234	0.152	0.175	0.111	0.152
MC-DE-30	P	0.865	0.984	0.185	0.287	0.301	-0.887
	S	1.0	1.0	-0.500	-0.500	0.500	-1.0
WS-DE-350	P	-0.085	0.063	0.185	0.188	0.127	0.187
	S	-0.157	0.009	0.172	0.194	0.142	0.142
RG-DE-65	P	0.800	0.558	-0.572	0.485	-0.166	0.233
	S	0.800	0.800	-0.8	0.399	0.200	0.200

Bibliography

[1] Ivan Klementiev, Alexandre; Titov. Crosslingual Distributed Representations of Words.