CLARIN-PL

Common Language Resources and
Technology Infrastructure

Wordnet-oriented recognition of derivational relations

WIKTOR WALENTYNOWICZ, MACIEJ PIASECKI

WROCŁAW UNIVERSITY OF SCIENCE AND TECHNOLOGY

Presentation plan

Introduction to the problem of derivation

Describe of derivation data from plWN

Our motivation

Representation of lexical units for classification

Classification tasks (and problems of it)

Conclusions, limitations and future work

Derivational relations in Polish

dom *a house*	domek *a little house*
lut *a solder (noun)*	lutować *solder (verb)*
kolarz *a cyclist*	kolarka *a female cyclist*

Our dataset

Lexical units from plWordNet 4.2

77122 pairs of (single word) lexical units linked by derivational relations

14 coarse-grained relations types

44 fine-grained relations types

Motivation

Support for linguistic research

Semi-automated verification of resources quality

Semi-automated extension of resources

Help in paraphrase task

Identifing meanings of unknown words

Different ways of representing a lexical unit

Bag of Characters vector

For each occurrence of a character in a word, the value of the vector at the index corresponding to that letter is increased by one.

In the case of Polish, the length of the vector is 35 -- 32 characters from the Polish alphabet plus the characters Q, V, X preserved for words borrowed from other languages.

We know exactly what a feature represents.

https://www.petplace.com/article/cats/pet-care/small-or-little-cats-name-ideas-for-small-or-little-cats/

Word form: kotek (*little cat*)

a	...	e	...	k	l	m	n	o	...	t	...	z
0	...	1	...	2	0	0	0	1	...	1	...	0

We know exactly what the features in the vector represent.

fastText vector

Vectors for a given word are created based on a learner model unsupervised on a large corpus.

The size of the vector depends on the parameters of the model with which it is generated.

Word form: kotek (*little cat*)

0	1	2	...	119	120	121	122	123	...	297	298	299
-0.161	-0.459	0.106	...	0.037	-0.114	0.030	0.121	0.397	...	-0.202	0.067	0.041

We do not know what specific feature values represent…

…but we know that the differences in the vectors should represent some relationship between the words.

fastText vector

Word form: kotek (*little cat*)

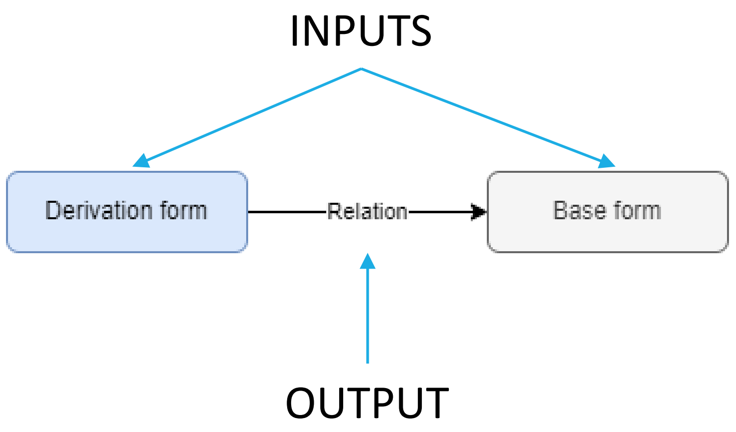
0	1	2	...	119	120	121	122	123	...	297	298	299
-0.161	-0.459	0.106	...	0.037	-0.114	0.030	0.121	0.397	...	-0.202	0.067	0.041

$$V_{kotek}^{w} = \frac{\sum\{V_{kot}, V_{ote}, V_{tek}, V_{kote}, \dots, V_{kotek}\}}{N}$$

N is the number of known n-grams for the vectorised word.

The orthographic form of a word affects the fastText vector, more so than is the case in Word2Vec vectors.

Different ways of representing learning examples



INPUTS
Derivation form
Relation
Base form
OUTPUT

Different ways of representing learning examples

DIFFERENTIAL VECTOR

It only represents the difference between the derivational form and the base form.

3-WAY VECTOR

Includes a full representation of the lexical units involved in the relation.

Vector combinations

Only BoC vectors

Only fastText vectors

Combined
fastText for words
BoC for differentia vector

Classification of relations

Type of classificators

Decision Tree

Random Forest

Multilayer Perctepron

Type of classificators

Decision Tree	Random Forest	Multilayer Perctepron

Representations based on the BoC, were tested on all three types of classifiers, fastText-based representations only on MLP.

Experiments setup

We used the implementation of classifiers from the scikit-learn package

The dataset was divided into 5 lexically separable parts (with respect to the base form)

The MLP network after the input layer had a linear layer of size 100

All classifiers, were multi-class classifiers

Problems from the point of view of the classification task

The dataset is highly unbalanced. The smallest class contains 300 examples, while the most numerous class contains over 13,000.

We decided on a multi-class classifier because it is difficult to generate negative examples in teaching a binary classifier.

Experiments results

	BoC Diff DT	BoC Diff RF	BoC Diff MLP	BoC 3-way DT	BoC 3-way RF	BoC 3-way MLP
F-1 Score	0,826	0,828	0,826	0,794	0,818	0,822
St. Dev.	0,005	0,004	0,005	0,005	0,004	0,004
	FT Diff 100	**FT 3-way 100**	**FT Diff 300**	**FT 3-way 300**	**COMB 100**	**COMB 300**
F-1 Score	0,816	0,830	0,828	0,842	0,830	0,842
St. Dev.	0,005	0,000	0,004	0,004	0,000	0,004

BoC – bag of characters, FT – fastText, COMB – combined vectors from BoC & FT
Diff – differentia vector representation, 3-way – 3-way (concatanation) vector representation
DT – decision tree, RF – random forest, MLP – multilayer perceptron

Conclusions & limitations

Classifiers using embedding vectors from fastText perform better, but this is a difference of about 1.5 percentage points.

The BoC representation is much more informative for humans and allows comparison of classifier behaviour with linguistic rules.

The similar results of the simple and sophisticated vectorisation of the examples, may point to a problem with the learning data for the task as a currently limiting factor in further progress.

Future work

We want to extend our research to other languages for which information on derivational relations is available.

We want to attempt to prepare derivational examples in the corpus so that we can transfer this task to the contextual space.

Thanks for your attention!

Contact by e-mail: wiktor.walentynowicz@pwr.edu.pl