Multilingual Modal Sense Classification Using a Convolutional Neural Network

Ana Marasović, Anette Frank

Research Training Group AIPHES, Department of Computational Linguistics, Heidelberg University

Overview

- **Modal verbs** (MVs) are **ambiguous** between:
  - **epistemic** sense (possibility)
    He could be at home.
  - **deontic** sense (permission/obligation)
    You can enter now.
  - **dynamic** sense (capability)
    Only John can solve this problem.
- MVs are used to implicitly express sentiment

  Refugees ***may$_{(de)}$ not (are forbidden to)*** *cross the borders.* ⇒ *Writer has* ***negative sentiment*** *towards refugees crossing the borders.*
- MSC is special case of **word sense disambiguation (WSD)**
  - MVs have **restricted sense inventory**
  - MVs act as **operators which take a full proposition as an argument**

CNN for Modelling Sentences

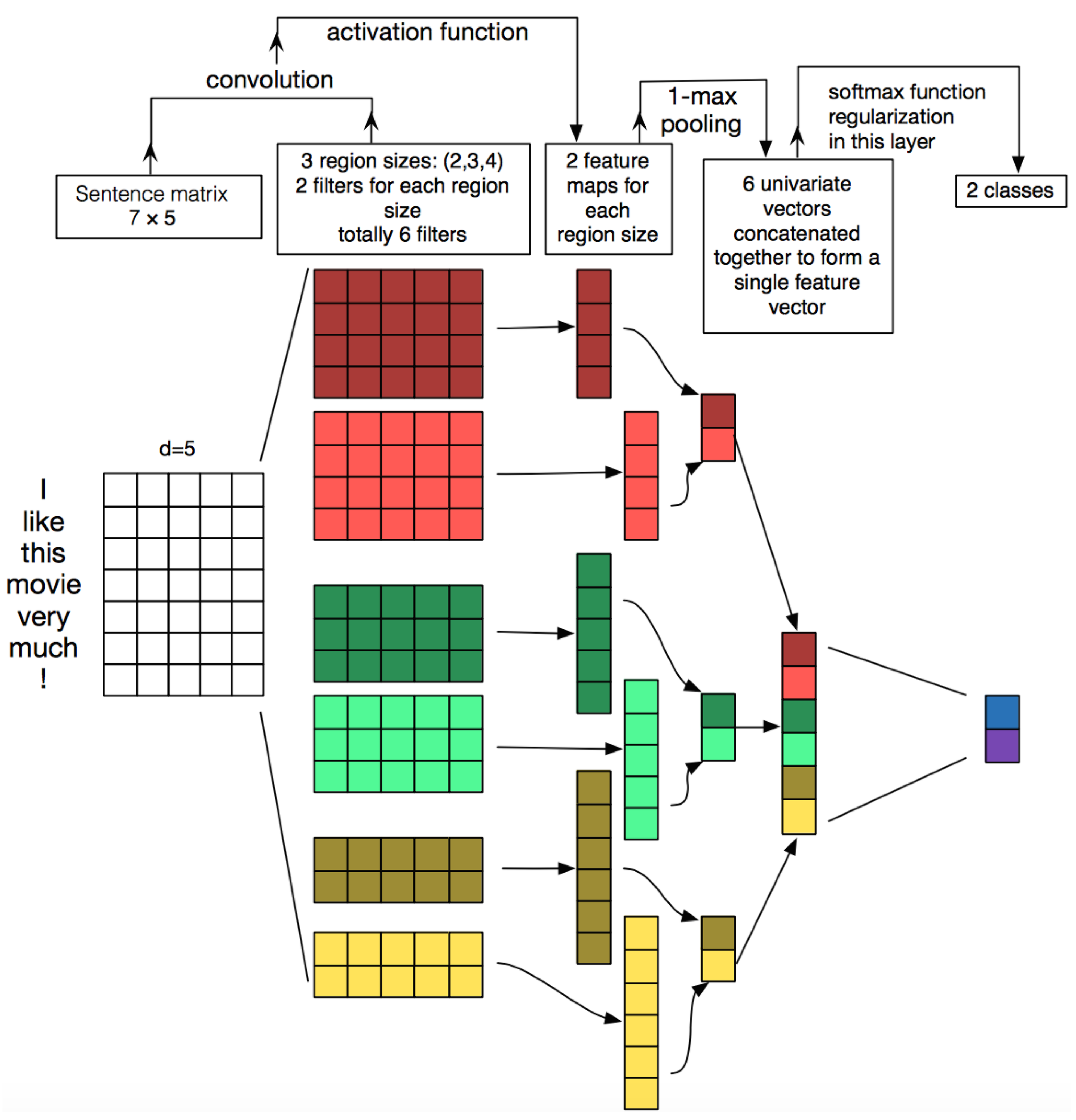

Figure 1: Convolutional neural network (CNN), Kim (2014).

Experimental Setup

Corpora
- **MPQA**: R&R's **small-scaled** manually annotated dataset with **strong sense bias**
- **EPOS$_E$, EPOS$_G$**:
  - a subset of EuroParl & OpenSubtitles corpora, heuristically tagged via the **cross-lingual sense projection** method of Z+

Model variations (for every modal verb)
- **CNN-E$_B$ / CNN-E$_U$**
  - 5-fold CV
  - train: **(un) balanced** MPQA + EPOS$_E$
  - test: MPQA
- **CNN-G**: train and test data from EPOS$_G$

Baselines
- random baseline BL$_{rand}$
- majority baseline BL$_{maj}$
- *MaxEnt* classifier from Z+
- one-layer neural network NN

Convolutional Neural Network for Modal Sense Classification

- **multilingual modal sense classification** using a **one-layer CNN architecture**
- **CNNs outperform strong baselines**, including **manually designed feature-based classifiers** and a plain NN
- **we identify known** and **previously untested semantic and linguistic features** from flexible window regions without syntactic pre-processing
- in a **standard WSD** task the CNN competes with a system using embeddings encoding richer information

MSC Results

	can	could	may	must	should	micro
BL$_{rand}$	33.33	33.33	50.00	50.00	50.00	41.49
MaxEnt	59.64	61.25	92.14	87.60	90.11	74.88
NN	56.01	55.42	90.00	75.24	88.68	69.74
CNN-E$_B$	**65.78**	**67.50**	**93.57**	**93.82**	**90.77**	**79.29**

	can	could	may	must	should	micro
BL$_{maj}$	69.92	65.00	93.57	94.32	90.81	80.18
MaxEnt	64.76	63.33	92.14	92.78	**91.48**	78.01
NN	67.29	66.08	**94.23**	86.37	90.96	77.93
CNN-E$_U$	**70.87**	**66.55**	93.49	**94.97**	90.59	**80.74**

Table 1: CV accuracies on MPQA with **balanced** (upper table) and **unbalanced** training (lower table).

	dürfen	können	müssen	sollen	micro
BL$_{rand}$	50.00	33.33	50.00	50.00	39.10
NN	77.73	43.32	73.88	50.25	57.69
CNN-G	**99.49**	**78.95**	**85.07**	**74.63**	**84.10**

Table 2: Accuracy on EPOS$_G$.

The CNN approach to MSC
- outperforms strong baselines
- **easily applicable to novel languages**
- reaches **high performance on German**
  - larger & perfectly balanced training data
  - Konjuktiv + tense ⇒ ep vs. de/dy

WSD Results

	SensEval-3
S$_{naive}$-prod	62.20
S-cosine	60.50
S-prod	64.30
S-raw	63.10
CNN	**66.50**

Table 3: WSD accuracy on *SensEval-3* dataset.

- *AutoExtend* encodes what is the target word, how many possible senses it has, and a sense embedding for each synset
- CNN without explicitly marking the target word outperforms *AutoExtend*

Feature Detectors

Filters are trained to **detect** semantic **features** found to be relevant in prior work, e.g. tense, aspectual classes, negation and semantic properties of verbs and phrases, as well as, **previously untested features.**

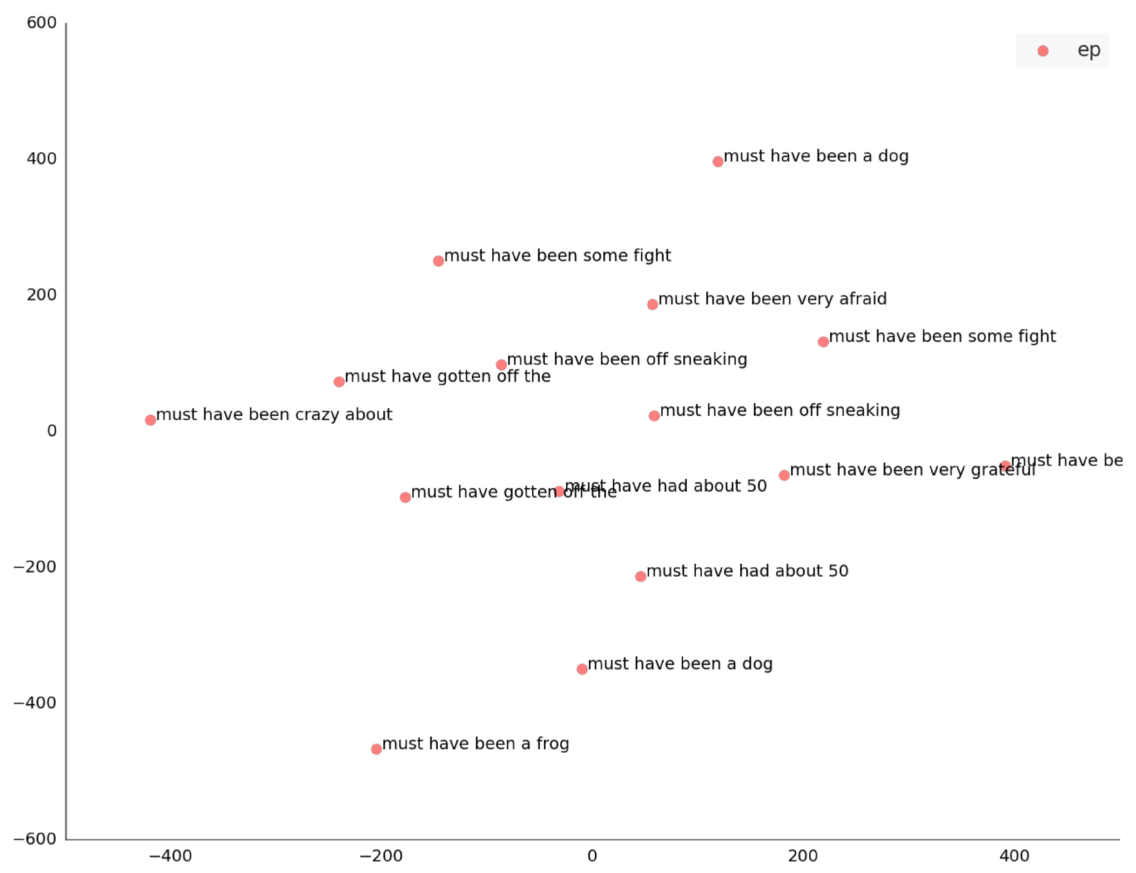

Figure 2: Feature detector that relates past reading of the embedded verb with epistemic sense for *must*.

features	sense	examples
past reading of the emb. verb	ep	you **must have been** out last night
non-past reading of the emb. verb	de	**we must take** further efforts
stative reading of the emb. verb	ep	**you must think** me a perfect fool
eventive reading of the emb. verb	de	**we must develop** a policy
passive construction	de	actual steps **must be taken**
negation	de	**we must not** fear
domain specific vocabulary	de	European parliament, present regulation, fisheries policy
telic clauses	de	to address these problems, to prevent both forum
discourse markers	de	but, and (then)

Figure 3: Other features detected for *must*, **including novel feature types.**

CNN for WSD

Data: *SensEval-3 lexical sample* dataset

Baseline: R&S's sense-specific embeddings

w . . . ambiguous word with k senses

c . . . centroid = sum of all w2v vectors of words in the sentence

$s^{(j)}$ . . . embedding of the j-th synset of w

S-cosine $= \langle \cos(c, s^{(1)}), \ldots, \cos(c, s^{(k)}) \rangle$

S-prod $= \langle c_1 s_1^{(1)}, \ldots, c_n s_n^{(1)}, \ldots, c_1 s_1^{(k)}, \ldots, c_n s_n^{(k)} \rangle$

S-raw $= \langle c_1, \ldots, c_n, \ldots, s_1^{(k)}, \ldots, s_n^{(k)} \rangle$

References

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the EMNLP*, pages 1746–1751, Doha.

Rothe, S. and Schütze, H. (2015). Autoextend: Extending word embeddings to embeddings for synsets and lexemes. In *Proceedings of ACL and IJCNLP*, pages 1793–1803, Beijing. Referred to as R&S.

Ruppenhofer, J. and Rehbein, I. (2012). Yes we can!? annotating the senses of english modal verbs. In *Proceedings of the LREC*, pages 24–26, Istanbul. Referred to as R&R.

Zhou, M., Frank, A., Friedrich, A., and Palmer, A. (2015). Semantically enriched models for modal sense classification. In *Proceedings of the LSDSem*, pages 44–53, Lisbon. Referred to as Z+.

Acknowledgements

This research was supported by the German Research Foundation as part of the Research Training Group "Adaptive Preparation of Information from Heterogeneous Sources" (AIPHES) under grant No. GRK 1994/1.