Copenhagen, task 1

Representation learning approach

Casper Hansen, Christian Hansen, Jakob Grue Simonsen, and Christina Lioma

UNIVERSITY OF COPENHAGEN

Check-worthiness objective

We consider the following requirements

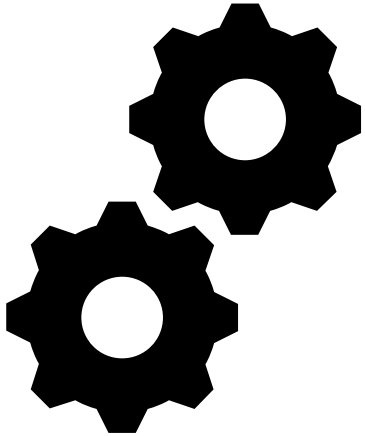
1. The sentence should be factually interesting,
2. and it should be possible to check

- (1): semantics and (2): semantics + structure
  - Semantics: word embedding
  - Structure: Simpler sentences through POS-tagging and through syntax dependency parsing

Speaker	Sentence
CLINTON:	I think my husband did a pretty good job in the 1990s.
CLINTON:	I think a lot about what worked and how we can make it work again...
TRUMP:	Well, he approved NAFTA... (check-worthy)

Related work

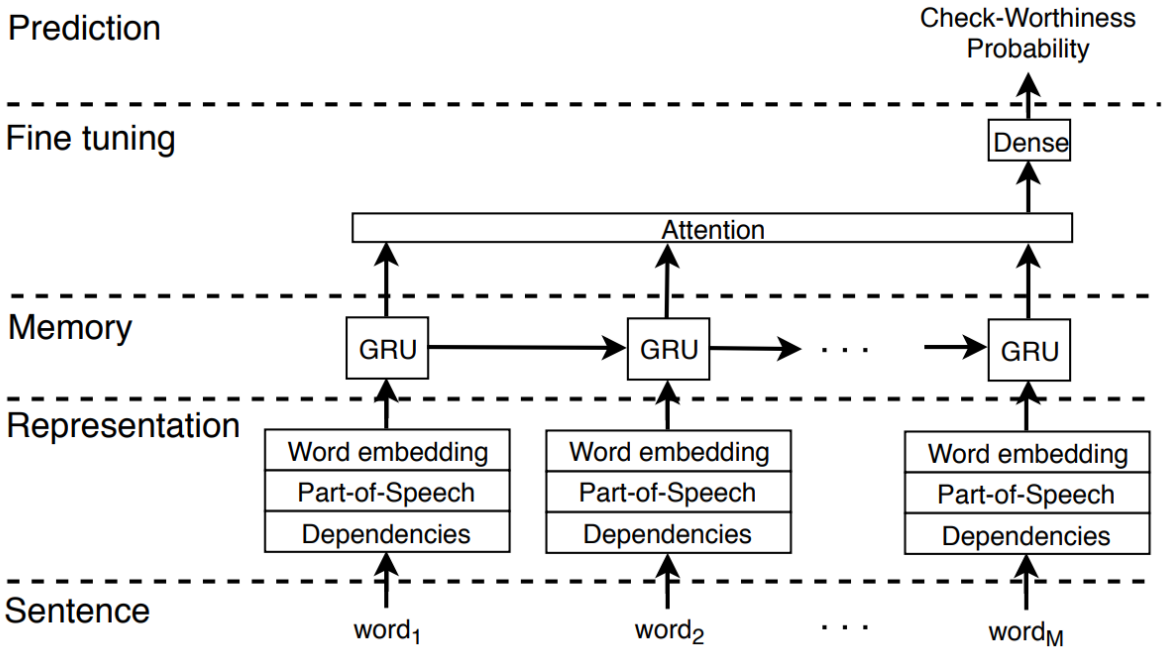
- Existing approaches most often use engineered features such as
  - bag-of-words (e.g. of tokens and POS tags)
  - Sentiment
  - Lexicon based features
  - Named entities
  - Length
  - Sentence position in debate segment
  - Speaker meta data
  - Topic modelling
  - Etc…



- Sentence level and contextual models

Representation learning approach

- Why not learn the features?
- Tokenization and lowercasing
- Word representations:
  - Fixed word embedding (word2vec, google news)
  - Syntax dependency parsing
  - POS tag



Experiments

- Comparison against two state-of-the-art systems: ClaimBuster [2] and contextual model by Gencheva et al. [1]
- We evaluate our sentence-level model with and without combining it with the contextual baseline model
- Experiment 1: debatewise comparison
- Experiment 2: speakerwise comparison

[1] Gencheva, P., Nakov, P., Màrquez, L., Barrón-Cedeño, A., & Koychev, I. (2017). **A context-aware approach for detecting worth-checking claims in political debates**. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017* (pp. 267-276).

[2] Hassan, N., Arslan, F., Li, C., & Tremayne, M. (2017). **Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster**. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1803-1812). ACM.

Debatewise comparison

- We compare models via **MAP**

Model	1st-president	2nd-president	vice-president	Average
ClaimBuster [2]	0.159	0.109	0.043	0.104
Gencheva et al. [1]	0.193	0.114	0.039	0.115
Our	0.109	0.206	**0.105**	0.140
Our + Gencheva et al. [1]	**0.225**	**0.207**	0.043	**0.158**

- Our approach performed better than baselines
- Combined approach outperformed all

Speakerwise comparison

Model	Trump	Clinton	Pence	Kaine	Average
ClaimBuster [2]	0.128	0.139	0.074	0.041	0.095
Gencheva et al. [1]	0.143	**0.201**	0.057	0.039	0.110
Our	0.225	0.077	**0.169**	**0.086**	**0.139**
Our + Gencheva et al. [1]	**0.245**	0.149	0.054	0.050	0.125

- Combined approach worse than our approach alone, but still better than baselines
- Small training set

Competition results

- Primary: combined approach
- Contrastive: our approach alone
- Contrastive: Best scores on all measures in competition (e.g. +30% better MAP)

	MAP	MRR	MR-P	MP@1	MP@3	MP@5	MP@10	MP@20	MP@50
Prise de Fer [25]									
primary	**.1332**$_{(1)}$	**.4965**$_{(1)}$	**.1352**$_{(1)}$	**.4286**$_{(1)}$	**.2857**$_{(1)}$	.2000$_{(2)}$	.1429$_{(3)}$	**.1571**$_{(1)}$	.1200$_{(2)}$
cont. 1	.1366	.5246	.1475	.4286	.2857	.2286	.1571	.1714	.1229
cont. 2	.1317	.4139	.1523	.2857	.1905	.1714	.1571	.1571	.1429
Copenhagen [11]									
primary	.1152$_{(2)}$	.3159$_{(5)}$	.1100$_{(5)}$	.1429$_{(3)}$	.1429$_{(4)}$	.1143$_{(3)}$	.1286$_{(4)}$	.1286$_{(2)}$	**.1257**$_{(1)}$
cont. 1	.1810	.6224	.1875	.5714	.4286	.3143	.2571	.2357	.1514
UPV–INAOE–Autoritas [9]									
primary	.1130$_{(3)}$	.4615$_{(2)}$	.1315$_{(2)}$	.2857$_{(2)}$	.2381$_{(2)}$	**.3143**$_{(1)}$	**.2286**$_{(1)}$	.1214$_{(3)}$	.0886$_{(4)}$
cont. 1	.1232	.3451	.1022	.1429	.2857	.2286	.1429	.1143	.0771
cont. 2	.1253	.5535	.0849	.4286	.4286	.2571	.1429	.1286	.0771

Conclusions

- Neural networks are able to learn a sentence representation at a similar or better level than engineered features, depending on the evaluation data.
- Differences between debates and speeches naturally influence model performance
- It could be interesting to expand the model with context by incorporating neighboring sentences.

Thanks!

Questions?