Detecting state of aggression in sentences using CNN

Denis Gordeev,
Scient. Adv. Rodmonga Potapova
Moscow State Linguistic University

Contents

- Previous research
- Model
- Data
- Results
- Conclusion

Introduction

Cyberbullying
Internet aggression
Prosecuted according to the Russian criminal code

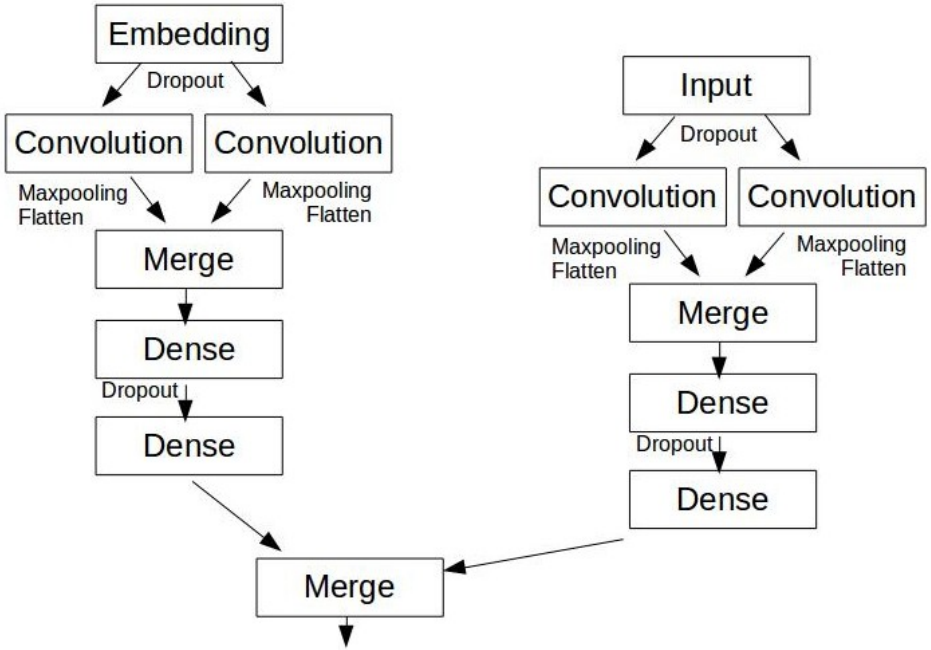
Related works

- L. Komalova, R. Potapova «On principles of annotated databases of the semantic field “Aggression”», SPECOM 2014
- Yoon Kim «Convolutional Neural Networks for Sentence Classification», 2014

Model

github.com/alexander-rakhlin/

Filters — 3 (instead of 100)
Batch 32



Data

Corpora	Avg. sentence length	Dataset (sent.)	Voc. size	Voc. In w2v model
Movie reviews	20	10662	18765	17121
SVAggr. (eng.) 4chan.org	19	19732	30765	28690
SVAggr. (rus.) 2ch.hk	13	5101	10030	9689

Results

Classifier	MR (%)	Verb. Aggr. (eng.)	Verb. Aggr. (rus.)
Random Forest	58.39	88.4	59.13
CNN-non-static	81.1	81.39	66.68
CNN (POS)	80.9	81.17	62.37
CNN-non-static, CNN (POS) combined	81	81.22	64.53

Conclusion

- The results of the POS-only classifier are surprisingly good. However, it requires further investigation. Moreover, merging the POS-classifier with the embeddings-based network does not improve the overall results.
- The results for the Russian language are much worse, than for English. So probably the system needs tweaking and including morphological and syntactic information

Thank you for your attention