ENTITY-LEVEL CLASSIFICATION OF ADVERSE DRUG REACTIONS: A COMPARISON OF NEURAL NETWORK MODELS

Ilseyar Alimova
Elena Tutubalina
Kazan Federal University

22 November 2018

Introduction

- Detection of new adverse drug reactions (ADR) in the post-approval period is on of the main task of pharmacology
- ADRs identified only post-marketing are a significant cause of morbidity and mortality
- Detection of new ADRs is also intended to drug reprofiling
- Unstructured texts are a promising source for information about adverse drug reactions

Previous approaches

- Based on classical machine learning models
- Evaluated on a single corpus
- Extracted information from the mention itself and a small window of words

Examples

non-ADR: "He was **unable to sleep** last night because of pain"

ADR: "Became unable to walk without a cane, **unable to sleep**, kidney problems"

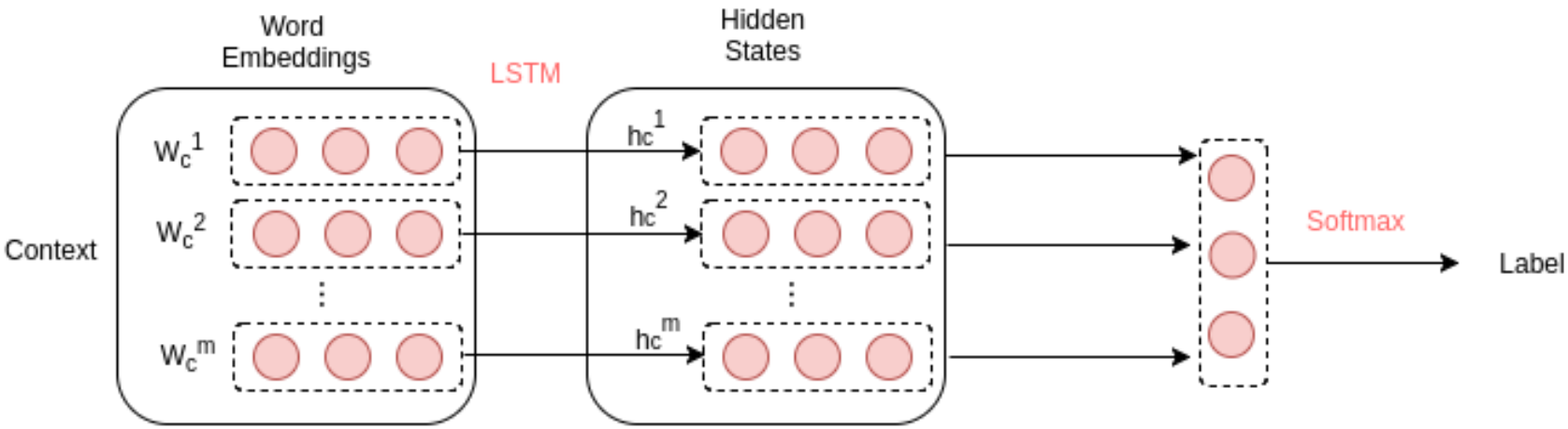
Our approach

- Based on Neural Network models
- Use as a context the whole sentence
- Evaluated on a different data sets
- Apply word embeddings trained on texts about health from social media

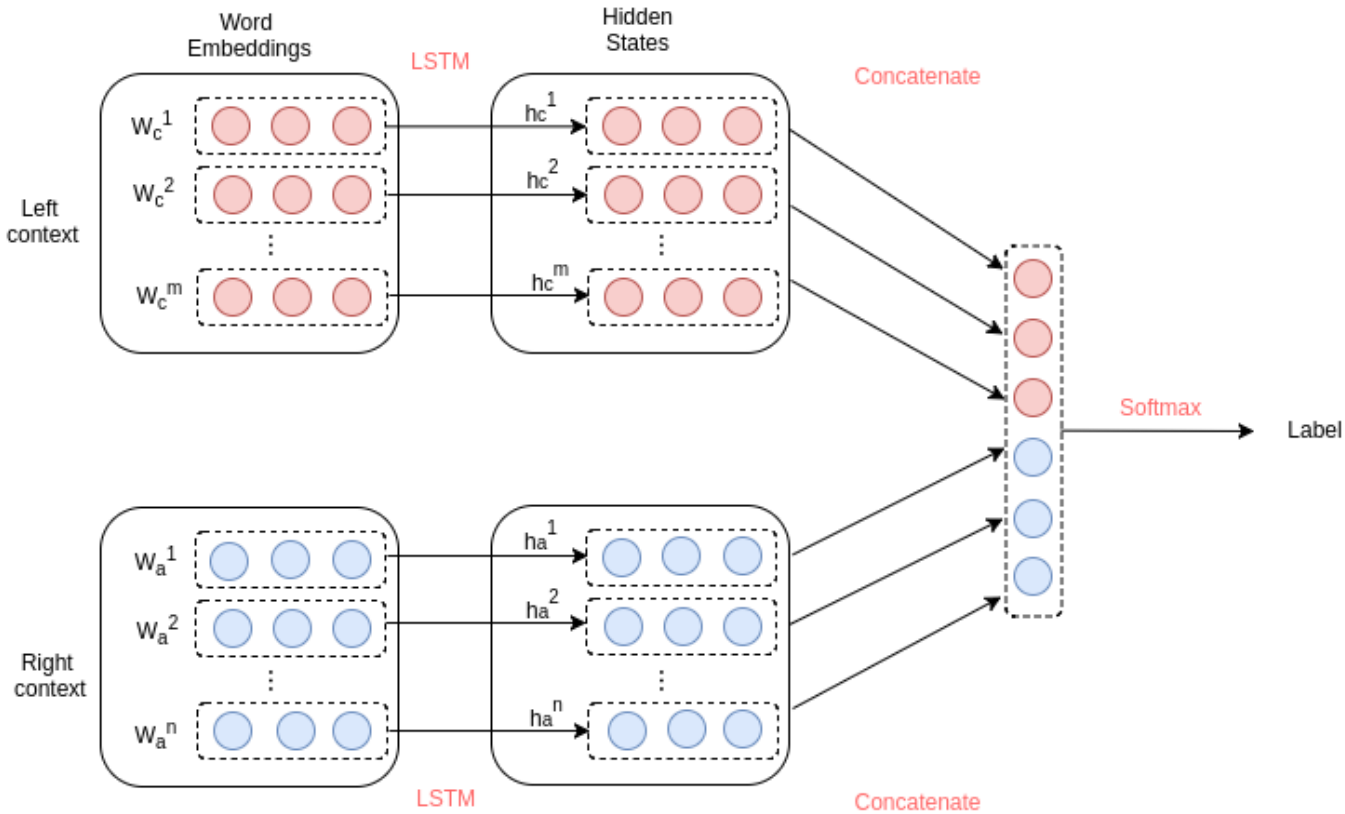
Neural Network Models

- Long Short Term Memory (LSTM)
- Target Dependency LSTM (TD_LSTM)
- Interactive Attention Network (IAN)
- Deep Memory Network (MemNet)
- Recurrent Attention Network on Memory (RAM)

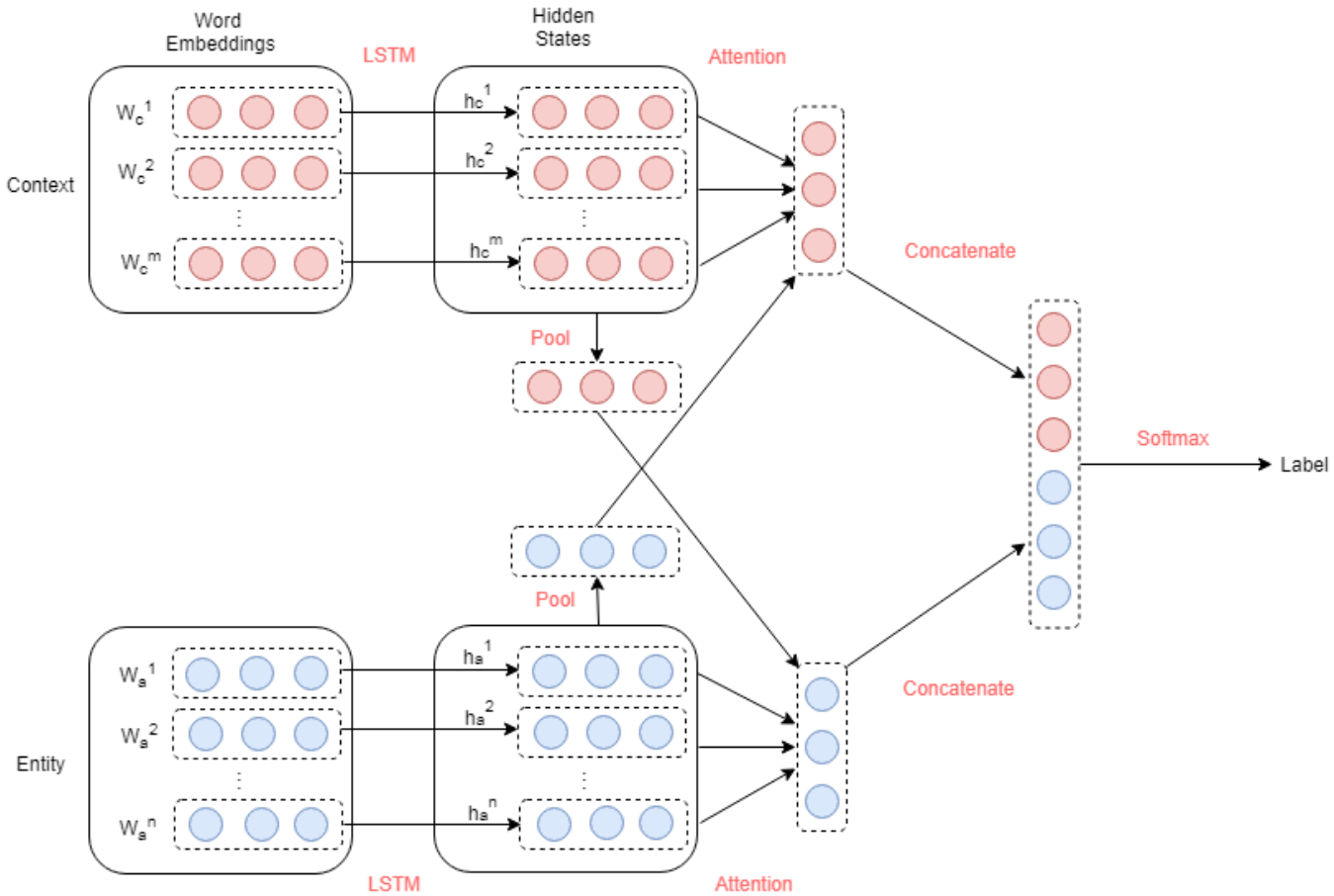
Long Short Term Memory (LSTM)



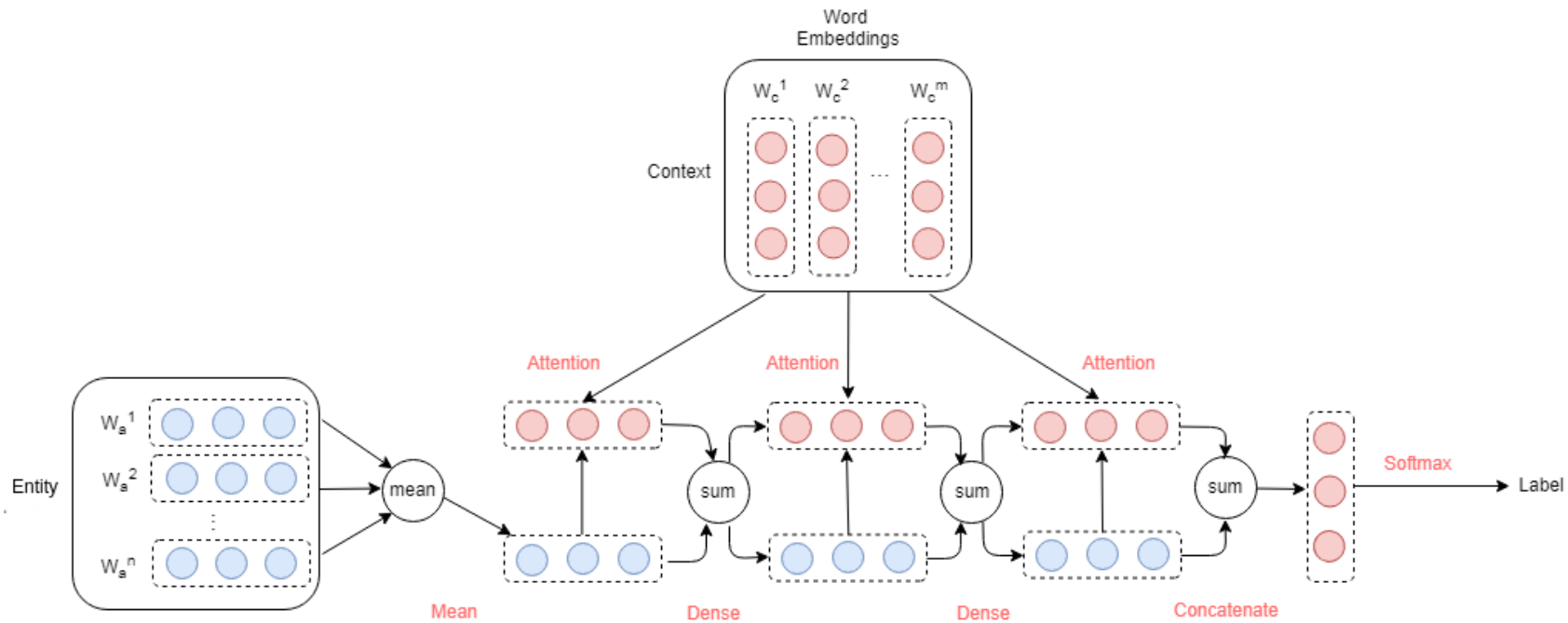
Target Dependency LSTM (TD_LSTM)



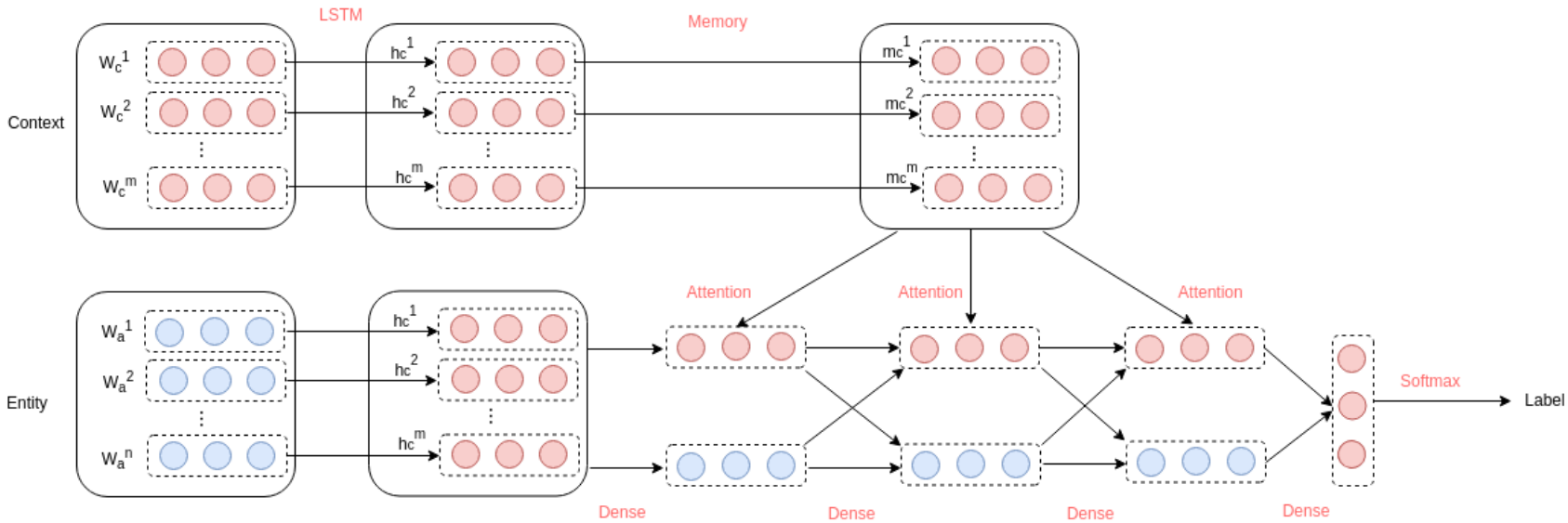
Interactive Attention Network (IAN)



Deep Memory Network (MemNet)



Recurrent Attention Network on Memory (RAM)



Datasets

- CADEC - user reviews from Askapatient forum
- MADE - electronic health records
- TwiMed-Pubmed - abstracts from PubMed articles
- TwiMed-Twitter - tweets about drugs
- Twitter - tweets about drugs

Corpus	Documents	ADR	non-ADR	Max sentence length
CADEC	1231	5770	550	236
MADE	876	1506	37077	173
TwiMed-Pubmed	1000	264	983	150
TwiMed-Twitter	637	329	308	42
Twitter	645	569	76	37

Word embeddings

- Trained on 2.5 million health-related reviews
- Contains 254 765 543 tokens
- Obtained with Continuous Bag of Words model
- Adopted from [Miftahutdinov et al., 2017]

Corpus	CADEC	MADE	Twimed-Pubmed	Twimed-Twitter	Twitter
Coverage	93.5%	62.5%	76.4%	81.2%	80.4%

Baseline Method

SVM with linear kernel and features:

- Bag of Words
- Part-of-speech tag
- Sentiment
- Cluster based
- Semantic types from Unified Medical Language System

Results of macro F-measure for both classes

	CADEC	Twitter	MADE	Twimed-Twitter	Twimed-Pubmed
SVM	.802	.749	.772	.758	.834
LSTM	.784	.613	.771	.700	.839
TD_LSTM	.772	.758	.750	.730	.709
MemNet	.758	.763	.760	.795	.811
RAM	.734	**.834**	.761	.780	.789
IAN	**.815**	.794	**.786**	**.819**	**.874**

- RAM outperformed other models on Twitter corpus
- IAN outperformed other methods on other four corpora
- the most significant increase was obtained on Twitter, Twimed-Pubmed and Twimed-Twitter corpora

Conclusion

- The potential of neural networks for aspect based sentiment classification to the task of ADR classification was explored
- RAM shows the best result for Twitter corpora and IAN shows the best results for other four corpora
- In the future: provide a detailed analysis of various kinds of errors and conduct experiments on Russian corpora

Acknowledgements

This work was supported by the Russian Science Foundation Grant.

Our team:

- Kazan Federal University
  - Elena Tutubalina
  - Timur Madzhidov
  - Zulfat Miftahutdinov
  - Valery Solovyev
  - Ramil Nugmanov
- NRC Kurchatov Institute
  - Alexander Sboev
  - Roman Rybka
  - Vladimir Gudovskyh
- Neuromation Company
  - Sergey Nikolenko
- Sechenov University
  - Sanna Sboeva

ENTITY-LEVEL CLASSIFICATION OF ADVERSE DRUG REACTIONS: A COMPARISON OF NEURAL NETWORK MODELS

Ilseyar Alimova
Elena Tutubalina
Kazan Federal University

22 November 2018