NATIONAL RESEARCH CENTER "KURCHATOV INSTITUTE"

EVALUATION OF MACHINE LEARNING METHODS FOR RELATION EXTRACTION BETWEEN DRUG ADVERSE EFFECTS AND MEDICATIONS IN RUSSIAN TEXTS OF INTERNET USER REVIEWS

Speaker:
Selivanov A.A.

Authors:
Sboev A.G.
Rybka R.B.
Moloshnikov I.A.
Rylkov G.V.

Relevance

Automatic extraction of medication adverse effects allows to expand information base for pharmacovigilance, that leads, on the one hand, to better quality control, on the other hand, allows successful reprofiling.

This work includes results on Russian Drug Review Corpus for the task of relation extraction between entities of the types ADR and Drugname.

Task formulation

Review containing medication and hypothetic adverse effects:

Арипризол долгое время вызывал тошноту и не помог против моих панических атак вызванных другим препаратом

The task is to determine if there is a relation between the pair of entities:

Арипризол + тошноту → есть связь

Арипризол + панических атак → нет связи

Used dataset

We used **Russian Drug Review Corpus** – a dataset of user reviews on medications, where experts manually annotated adverse drug reactions and medication names, as well as relations between them:

- **4289** pairs of entities with relation;
- **1132** pairs of entities without relation.

Experiment setup

- Stratified split on train (80%) and test (20%) sets;
- Relations from single review could be in one set only;
- Hyperparameters of basic methods were chosen on test set, reason: we want to obtain maximum theoretical accuracy as a baseline to overcome;
- Hyperparameters of language model fine-tuning were chosen on validation part of a train set to estimate model accuracy in circumstances close to real;
- Language model used early stopping and learning rate decay (which have positive effect on model accuracy);
- This work has been carried out using computing resources of the federal collective usage center Complex for Simulation and Data Processing for Mega-science Facilities at NRC “Kurchatov Institute”, http://ckp.nrcki.ru/.

Methods

The following methods were used to solve relation extraction task:

- Random stratified generation;

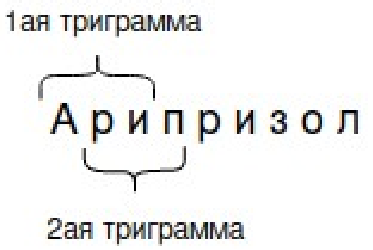
- Logistic regression;
- Support vector machine;
- Naïve Bayes Model;
- Gradient Boosting.

- Language model RuDRBERT (multilingual BERT with additional training on 1.4M of user-generated texts on health&medicine thematic).

Feature engineering (basic methods)

Basic machine learning methods used the following text data representation:

- Text only of the entities in current pair;
- Character n-grams;
  - n is a hyperparameter, could be interval;
- TF-IDF of the character n-grams
  - frequency filter applied, values of the filter are also hyperparameters).



- Hypothesis: such a little data (only text of the entities) is still enough to achieve competitive accuracy for the task.

Feature engineering (language model)

Experiments included the following sets of options for text representation:

1) Text part:
   - Whole text;
   - Text of an entities pair;
   - A pair of entities, separation token, text between entities;
   - A pair of entities, separation token, whole text.

2) Entities screening:
   - No screening.
   - Only entities of target pair;
   - All entities;

Example:

Yesterday I get a shot of **Sputnik-V**, now I feel a little **fever**, but I heard it’s still better than **EpiVAC**.

Yesterday I get a shot of **[T_MED]Sputnik-V[/T_MED]**, now I feel a little **[T_ADR]fever[/T_ADR]**, but I heard it’s still better than EpiVAC.

Yesterday I get a shot of **[T_MED]Sputnik-V[/T_MED]**, now I feel a little **[T_ADR]fever[/T_ADR]**, but I heard it’s still better than **[1_MED]EpiVAC[/1_MED]**.

3) Adding special tokens to model vocabulary.

Results (basic methods)

The table shows results with optimal hyperparameters

Model	Macro-averaged f1
Stratified random class (dummy model) (average on 100 executions)	0.54
Logistic regression	0.72
Support vector machine	0.74
Multinomial Naive Bayes Model	**0.75**
Gradient Boosting	0.74

Results (language model) (1 of 2)

Group	Feature	Macro-averaged f1-score
Without fine-tuning, no special tokens	All entities screened	0.44
	Target entities screened	0.45
Without fine-tuning, with special tokens	All entities screened	0.44
	Target entities screened	0.44
With fine-tuning, No special tokens	All entities screened	0.47
	Target entities screened	**0.68**
With fine-tuning, with special tokens	All entities screened	0.49
	Target entitites screened	**0.67**

Results (language model) (2 of 2)

Hyperparameters:

- Fine-tuning;
- Target entities screening;
- Early stopping;
- Maximum tokens: 512;
- Learning rate: 0.00001;
- Batch size: 32;
- Epochs: 10;
- Learning rate decay.

Text representation	Macro-averaged f1-score
Whole text	0.68
text of target entities only	0.73
text of target entities + text between them	0.77
text of target entities + whole text	**0.88**

Conclusion

- Baseline obtained for pharm entities relation extraction between medication and adverse effects, which is **54%** (based on stratified random classification dummy algorithm);
- It is shown that basic machine learning algorithms that uses target entities character n-grams tf-idf with frequency filtering allows to achieve accuracy equal to **75%** (Multinomial Naive Bayes), which is **21%** higher than random class prediction. We could think of it as a maximum for such task formulation due to hyperparameter search on a test set and in further research this value could be used as sophisticated baseline;
- A set of computational experiments were carried out using language model, pre-trained for medical domain, obviously fine-tuning allows to get higher accuracy, surprisingly excluding special tokens of entity screening from the model vocabulary doesn“t affect accuracy much;
- It is shown that the best accuracy achieved with the text representation, that includes target entity text, separation token and the whole text. Hypothetically, it allows to form representation of each entity;
- As a result, best achieved accuracy (macro-averaged f1) with language model is **88%**, which is **34%** higher than random class prediction, **13%** higher than basic machine learning models with hyperparameters tuned on test set;
- This results could be used as an important step in analysis of medical texts in Russian, in particular in extraction of relations between medications and their adverse effects.

Thank you for attention