Automatic Recognition of Unfair Clauses in Polish Consumer Contracts

Michalina Skibicka, MA; supervision: Lukasz Gorski, Ph. D.

Interdisciplinary Centre for Mathematical and Computer Modelling (ICM), University of Warsaw

Introduction

The project is built around the **register** of unfair contractual clauses by **UOKiK**, Polish Office of Competition and Consumer Protection: https://decyzje.uokik.gov.pl/ used as a corpus base for analysis.
The novelty of the project lies in the fact that:

- the matter has **not** been **extensively researched before**, in particular - no substantial research for automatic analysis of Polish legal texts found as far (Humphreys et al. 2015, Lippi et al. 2017);
- initial commercial research shows **niche in the market**: within a sample of ca. 50 subjects recruited among legal professionals, corporate and individual customers **over 80%** expressed a decisive need for some form of a tool for automatic recognition legal documents, 13% were favorable to its use and ca. **70%** declared the willingness to pay for such a solution (online questionnaire and in-person interviews, not yet published);
- **specific linguistic composition** of legal texts (high formality and conventionalisation level, high density of noun phrases, high average sentence length, ubiquitous paraphrasing within a limited intention-base) offers an interesting opportunity for the use of both **classical NLP methods** and **ML-based** solutions.

Initial dataset analysis

The dataset is composed of **7092 clauses**, with large sentence length variation, the longest being 51 tokens.

First task: **basic NLP processing** of the text with the use of available Polish-relevant tools:

- Korpusomat: http://korpusomat.pl/ - Polish BNLP processing tool outputting an XML file with lemmatization, POS tagging and morphosyntactic to be used in further proceedings, i.e. rule-writing.

Figure 1: Kormusomat XML output

- TermoPL: https://ws.clarin-pl.eu/termopl.shtml - terminology extraction tool compatible with Korpusomat language information, to be used as a Named Entity source for filtering and rule-writing thanks to the possibility of lemma-based extraction of multi-token expressions (Mykowiecka et al. 2016)

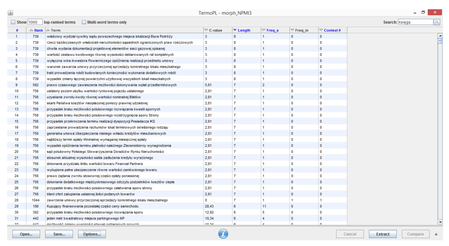

Figure 2: TermoPL output

- Topic: https://ws.clarin-pl.eu/topic.shtml - implementation of topic modelling for Polish in Mallet:

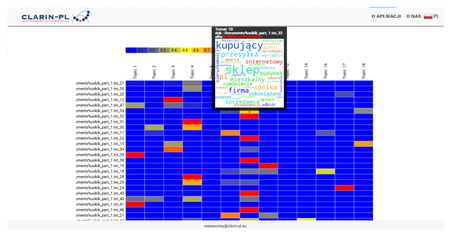

Figure 3: Topic Mallet output

and Gensim:

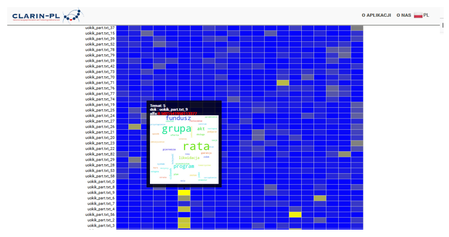

Figure 4: Topic Gensim output

intended to be used as a first stage of automatic clause recognition in the project - as far the register contains only clauses relevant to a number of topics concerning **consumer law** cases, for example with **no labour law** or **employment contracts**. Topic offers an API access to a REST application, so some extent of integration within the project is possible (Piasecki et al. 2018)

Working with the data

1. Manual grouping and human labelling of clauses

Based on their abusivity, 6 groups were distinguished (Romanowski 2010; Act on competition 2007):
- contractual enforcement of the localization of a competent court (label: SAD),
- imposing an unfair contractual penalty (label: KARA),
- requiring an unjustified payment: keeping an already paid fee or requiring additional payment (label: ZATRZOPL and DODOPL),
- lack of transparency in stipulation of contractual conditions,
- abusive disproportion of rights and obligations of the parties,
- decontextualized clauses - impossible to analyze outside of full document meaning.

As far: finished labeling of the first three groups:
- SAD: 505 clauses,
- KARA: 231 clauses,
- ZATRZOPL and DODOPL: 703 clauses,

all amounting to the total of 1439 clauses, being 20% of the corpus.
TBD: exploring the possibility of **auto-encoder** use in automatic clause grouping.

2. Classification of clauses

Done for SAD and KARA labels with **Sklearn** implementation of two classifiers: **Naive Bayes** and **LSVM** with and without the use of TF-IDF feature:

```
Multinomial naive Bayes accuracy = 0.9796195652173914
SVM accuracy = 0.9932065217391305
              precision    recall  f1-score   support

        KARA       0.95      0.99      0.97       231
         SAD       0.99      0.98      0.99       505

   micro avg       0.98      0.98      0.98       736
   macro avg       0.97      0.98      0.98       736
weighted avg       0.98      0.98      0.98       736

              precision    recall  f1-score   support

        KARA       0.98      1.00      0.99       231
         SAD       1.00      0.99      1.00       505

   micro avg       0.99      0.99      0.99       736
   macro avg       0.99      0.99      0.99       736
weighted avg       0.99      0.99      0.99       736
```

Figure 5: NB and LSVM with TF-IDF results

Such results can be attributed to the lexical disparity between two groups of clauses and are expected to drop with the addition of the next less disparate groups of labels. Results slightly improved for NB and slightly dropped for LSVM without the use of TF-IDF feature:

```
Multinomial naive Bayes accuracy = 0.9823369565217391
SVM accuracy = 0.9918478260869565
              precision    recall  f1-score   support

        KARA       0.95      0.99      0.97       231
         SAD       1.00      0.98      0.99       505

   micro avg       0.98      0.98      0.98       736
   macro avg       0.98      0.98      0.98       736
weighted avg       0.98      0.98      0.98       736

              precision    recall  f1-score   support

        KARA       0.98      1.00      0.99       231
         SAD       1.00      0.99      0.99       505

   micro avg       0.99      0.99      0.99       736
   macro avg       0.99      0.99      0.99       736
weighted avg       0.99      0.99      0.99       736
```

Figure 6: NB and LSVM without TF-IDF results

TBD: implementation of **Tensorflow-based** classifiers; use of manually written implementation of TF-IDF in a customized classifier for comparison of methods.

3. Human annotation and extraction rules

Trials done for the less lexically repetitive group ZATRZOPL and DODOPL.
Rules: CoreNLP TokensRegex demo: https://stanfordnlp.github.io/CoreNLP/

At present: surface-level work as no support for Polish. **TBD**: introduction of Polish Korpusomat XML file as input.

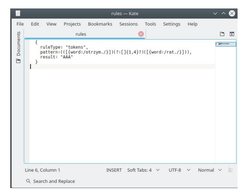

Figure 7: TokensRegex rule

Issue: limited no. of expressions supported. **TBD**: manual introduction of **lookaround** expressions. If impossible: **SemGrex demo** trials.

Figure 8: TokensRegex rule output on mini-corpus

Additionally: initial work on a customized **synonyms dictionary** for this group: oplaty, platnosci, obciazy...

Annotation: trials done manually, **TDB**: use of **BRAT** annotation tool: http://brat.nlplab.org/index.html

To be used as a classifier input for less lexically repetitive clauses. Item annotated: core abusivity notion. **TBD**: classifier options exploration.

Future developments

- extension of the task to **clause extraction**: document corpus needed from UOKiK;
- **recitals** mapping;
- **bag-of-words** model introduction;
- **word2vec** model introduction (including **customized** legal **conversion** from EurLex bases);
- **paraphrase detection** task for clause recognition;
- **i**ntroduction of relevant **similarity measures**: Jaccard, cosine, entropic, Pearson

References

- Act of 16 February 2007 on competition and consumer protection (Journal of Laws No 50, item 331)
- Das B., Chakraborty S. (2018) An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation
- Humphreys L. et al. (2015) Mapping Recitals to Normative Provisions in EU Legislation to Assist Legal Interpretation
- Lippi M., Palka P., Contissa G., Lagiola F., Micklitz H. W., Panagis Y., Sartor G., Torroni P. (2017) Automated Detection of Unfair Clauses in Online Consumer Contracts
- Mykowiecka A., Marciniak M, Rychlik P. (2016) TermoPLâĂŤ a Flexible Tool for Terminology Extraction
- Piasecki M., Walkowiak T., Eder M. (2018) Open Stylometric System WebSty: Integrated Language Processing, Analysis and Visualisation
- Romanowski M. (2010) W sprawie charakteru i skutkow abstrakcyjnej kontroli niedozwolonych postanowien wzorcow umownych stosowanych przez przedsiebiorce