Rhetorical Structure Theory as a Feature for Deception Detection in News Reports in the Russian Language

Why fake news?

"Post-truth" - word of the Year 2016 [Oxford Dictionaries]:
"relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief"

Why fake news?

- **We get information from different sources and should evaluate the reliability to avoid rumours, hoaxes and deceptive (fake) information in news reports.**

Sir Tim Berners-Lee, inventor of the worldwide web, included the spread of misinformation to the main threats to the web (2017).

Tim Cook, the head of Apple, declared that technology companies need to create some tools that help diminish the volume of fake news (2017).

Almar Latour, executive editor for The Wall Street Journal, hopes that the intrusion of fake news into the media ecosystem would remind of the extraordinary value of truth and quality journalism (2016).

Mark Zuckerberg, the chairman of Facebook, stated the need for technical systems to detect what people will flag as false before they do it themselves (2016).

Fake news detection as a problem

Human ability to detect misinformation score:

English: 0.54

Russian: 0.55 (later in the presentation)

Tools for automated deception detection and information verification, created for different languages (i.e. Russian), based on Natural Language Processing methods and models, are required in our society.

Possible applications: trends monitoring in social media, linguistic expertise, fact-checking tools for newsrooms and news aggregators etc.

Types of deceptive news

- serious fabrications;
- large-scale hoaxes;
- humorous fakes.

[Rubin V. L., Conroy N. J., Chen Y. (2015), Deception Detection for News: Three Types of Fakes]

-satire;
-extreme bias;
-conspiracy theory;
-rumour;
-state news in repressive states;
-junk science;
-fake news;
-clickbait;
-proceed with caution;
-political;
-credible.

[http://www.opensources.co/]

Features for deception detection in NLP

Lexics: part of speech, length of words, subjectivity terms, numbers and imperatives in headlines, frequency of affective words or action words from psycholinguistics lexicons (LIWC) etc: accuracy up to 0.77.
Syntax: patterns which help to distinguish types of arguments; rule categories from Probability Context Free Grammars: accuracy up to 0.91.
Lexics+syntax: different predicate types
Semantics: text coherency to similar texts
Pragmatics: pronouns with antecedents

Rhetorical structures 0.63 accuracy
[Rubin V. L., Conroy N. J., Chen Y. C. (2015), Towards News Verification: Deception Detection Methods for News Discourse]

Research objective

Objective is to reveal significant differences between structures of truthful news reports and deceptive ones, using RST relations as deception detection markers, based on the definite corpus:

- what the features should look like: are RST relation types' frequencies, relations' sequences important?
- estimate the impact of these features in detection: classify the texts, based on the RST relations labeling, and predict if news reports are truthful or deceptive.

Data collection principles

Lack of sources in Russian that contain verified samples of fake and truthful news (Factbanks, objective and impersonal fact checking websites).

The only way out in solving the problem was the reliance on the **presented facts, on the factuality.**

Data collection principles

The daily manual monitoring of news: 11 months (June 2015-April 2016).

Sources: Online media in Russian:

-well-known news agencies' websites and local or topic-based news portals;

-online newspapers from different countries (Russia, Ukraine, Armenia etc.).

Final data set consists of news reports dedicated to **38 different topics**, with equal number of truthful and deceptive news stories to each topic, and not more than 12 news reports about the same topic.

Each topic was **analyzed carefully** to define a fake part in the case and to avoid subjectivity and biased evaluation.

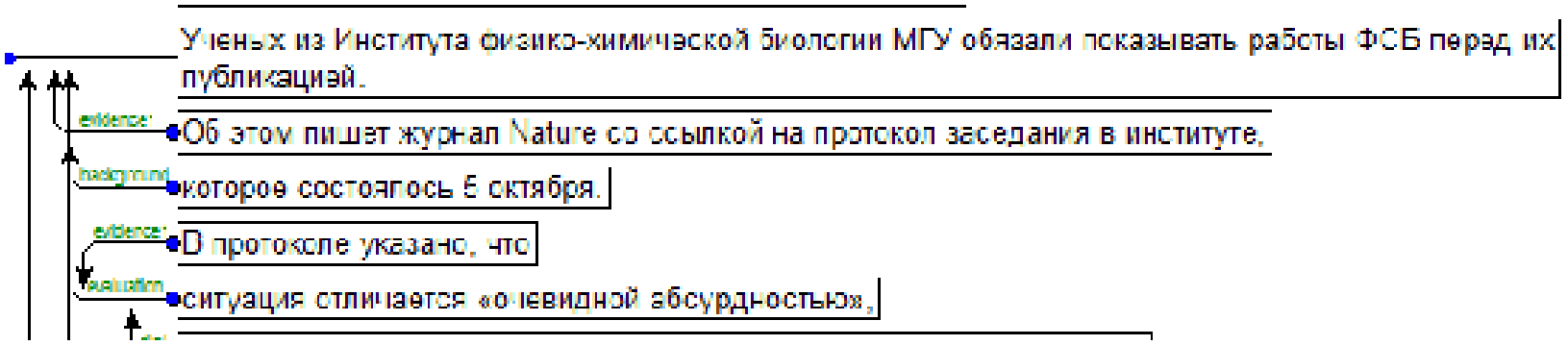
Corpus details

134 news reports, with average length 2700 symbols.
Average number of rhetorical relations in text is 17.43.
The whole number of rhetorical relations in corpus is 2340.
Clauses were taken as elementary discourse units.
There are no discourse parsers for Russian, that's why tagging and validation were made manually.

Annotation details

We used UAM CorpusTool for discourse-level annotation.

33 relation types: 'Circumstance', 'Reason', 'Evidence1', 'Evidence2', 'Evidence3', 'Evidence4', 'Contrast', 'Restatement', 'Disjunction', 'Unconditional', 'Sequence', 'Motivation', 'Summary', 'Comparison', 'Non-Volitional Cause', 'Antithesis', 'Volitional Cause', 'Non-Volitional Result', 'Joint', 'Elaboration', 'Background', 'Solution', 'Evaluation', 'Interpretation', 'Concession', 'Means', 'Conjunction', 'Volitional Result', 'Justify', 'Condition', 'Exemplify', 'Otherwise', 'Purpose'.



Inter-annotator agreement

- 2 annotators (66 reports and 68 reports).

Truthful and deceptive news reports about the same event were annotated by the same person.

- Discrepancies: Background/Sequence/Elaboration; Reason/Unvolitional Cause/Volitional Cause; Purpose/ Unvolitional Result/Volitional Result; Evaluation/Interpretation; Antithesis/Contrast; Elaboration/Justify/Restatement in quotations.
- Discussions.

2 steps of measuring Krippendorff's unitized alpha: 0.75 after the second step.

Main experiments

1th experiment
Baseline: lexics level: frequency of lemmas from a sentiment lexicon as a feature for each text: a list of 5000 sentiment words [Chetviorkin and Loukachevitch (2012), Extraction of Russian Sentiment Lexicon for Product Meta-Domain].
2th experiment
2.1. Model A: RST relation types frequencies
2.2. Model B: RST relation types frequencies +count of bigrams and trigrams
2.3. Model C: RST relation types frequencies + count of top 20 bigrams of RST types and top 20 trigrams of RST types

Two supervised learning methods for texts classification and machine learning: Support vector machines (SVMs) (linear and rbf kernels) and Random Forest, both with 10-fold cross-validation.

Additional experiment

The corpus was annotated manually to compare machine learning results, which are based on RST-features, with human asessments.

Results: 1 and 2 experiments

	Precision	Accuracy	Recall	F-measure
Support Vector Machines, rbf kernel, 10-fold cross-validation				
Baseline	0.38	0.42	0.54	0.42
Model A	0.54	0.53	0.51	0.51
Model B	0.60	0.55	0.52	0.50
Model C	0.65	0.61	0.56	0.57
Support Vector Machines, linear kernel, 10-fold cross-validation				
Baseline	0.23	0.37	0.49	0.31
Model A	0.64	0.65	0.65	0.63
Model B	0.64	0.60	0.48	0.53
Model C	0.62	0.59	0.60	0.59
Random Forest Classifier, 10-fold cross-validation				
Baseline	0.48	0.48	0.55	0.49
Model A	0.56	0.54	0.45	0.47
Model B	0.60	0.63	0.56	0.56
Model C	0.57	0.55	0.46	0.49

Results: additional experiment with human asessments

	Precision	Recall	F-measure
Scores for human asessments	0.55	0.46	0.50

Most significant features

The most significant features which influence on linear SVMs classification for model A are: 'Justify', 'Evidence3', 'Contrast', 'Evidence1', 'Volitional Cause', 'Comparison".

Relation type	p-value
Justify	**0.00018**
Evidence3	**0.02968**
Contrast	**0.03145**
Evidence1	**0.00209**
Volitional Cause	**0.03419**
Comparison	0.07858

Discussion-1

-The hypothesis is confirmed: there are differences between structures of truthful news reports and deceptive ones. The results for Russian (0.65) can be compared with the predictive power of the model for English (0.63).

-The model should be developed and modified, learned and tested on larger data collections with different topics.

-We should use a complex approach and combine this method with other linguistics and statistical methods.

-The guidelines for gathering a training corpus of obviously truthful/deceptive news should be improved.

Discussion-2

-The extrapolation of the existing model to all possible news reports in Russian, devoted to different topics, would be incorrect. But it could already be used in some cases as a preliminary filter for deceptive (fake) news detection.

-We tried to take into consideration RST-'trees'. It should be studied more deeply and intensively.

-The model is also restricted by the absence of automated discourse parser for Russian.

-The assignment of RST relations to news report could be connected with the subjectivity of annotators' interpretation. Manuals for tagging and by developing consensus-building procedures should be impoved.

Thank you for your attention!

Questions:
Dina Pisarevskaya
dinabpr@gmail.com

What news report is deceptive?

1) Вчера обескураженные жители Венеции обнаружили в своем канале кита, который заплыл туда из Атлантики. Глобальное потепление, видимо, сбило кита с “курса” и он вместо Северной Атлантики, через Гибралтар оказался в Средиземном море. Это второй случай, когда кит оказался у берегов Венеции. Первый раз это было во время Второй мировой войны. Тогда был голод. Кит был местными жителями убит и съеден. Сегодня не война и никто не собирался есть кита, но кит доставил не мало проблем. Его мощный хвост играючи потопил несколько лодок.

2) Вынесенная туша мёртвого кита, которая с начала ноября лежит на Средиземном побережье Франции, от скопления газов может взорваться в любую секунду. Французский телеканал BFM показал видеозапись с раздувающимся телом кита, которого вынесло на побережье Франции в начале ноября. Тушка выглядит как плотно накачанный воздушный шар, который вот-вот разорвётся. Дело в том, что кит уже начал разлагаться к тому моменту, как попал на берег. Территориально это место находится недалеко от Монпелье. Внутри кита из-за процессов гниения активно скапливаются газы, пишет The Local. На данный момент власти Франции бьют тревогу и пытаются скорее решить вопрос с огромной бомбой замедленного действия.