Deception Detection in News Reports in the Russian Language: Lexics and Discourse

Dina Pisarevskaya
dinabpr@gmail.com

Abstract

Different language markers can be used to reveal the differences between structures of truthful and deceptive (fake) news. Two experiments are held: the first one is based on lexics level markers, the second one on discourse level is based on rhetorical relations categories (frequencies). Corpus consists of 174 truthful and deceptive news stories in Russian. Support Vector Machines and Random Forest Classifier were used for text classification. The best results for lexical markers we got by using Support Vector Ma-chines with rbf kernel (f-measure 0.65). The model could be developed and be used as a preliminary filter for fake news detection.

1 Introduction

The research field of deception detection in news reports and automated fact checking arose in natural language processing (NLP) rather recently. It can be applied for linguistic expertise, fact checking tools for newsrooms and news aggregators, tools for users.

We get information from different sources and should evaluate the reliability to avoid rumours, hoaxes and deceptive (fake) information in news reports. The word 'post-truth' was chosen as the Oxford Dictionaries Word of the Year 2016 and points that objective facts can be less influential than appeals to emotion and personal belief. It regards political and other news of our 'post-truth era'. In the media community, key persons pay attention to the value of truth in journalism, to the necessity of fact checking, to the threat of fake news and to the need for technical systems that would help diminish the problem: Almar Latour (The Wall Street Journal, 2016), sir Tim Berners-Lee (worldwide web inventor, 2017), Tim Cook (Apple, 2017), and Mark Zuckerberg (Facebook, 2016).

There are three types of fake news: serious fabrications, large-scale hoaxes and humorous fakes (Rubin et al., 2015a). OpenSources (www.opensources.co) suggests more news types, fake news among them. They are understood as fabricated information, disseminated deceptive content, or grossly distorted actual news reports. This definition corresponds to serious fabrications.

In social media, people participate in the propagation of the news that they find interesting. Algorithmic ranking, social bubbles and group polarization may lead to intentional or unintentional viral spread of unreliable news. Big amounts of news reports with misinformation spread caused by political reasons in 2016 in the USA (presidential election) (Allcott and Gentzkow, 2017). For the Russian language the problem of fake news is already vital since 2014 (Russian-Ukrainian discourse).

2 Related Work

Data science companies, academics, media organizations are working on computational fact checking for English: on fake news detection and real-time detection algorithms. In 2016, Google gave funding to FactMata and Full Fact project to develop automated fact checking tools. FactMata's (UK) project is devoted to fact checking and claim validation by known statistical databases. The Full Fact (UK) is developing an automated fact checking helper, using the logic of question answering machines: facts from social media will be parsed, compared with curated known-true facts and determined as true or false. Tracer News system (0.84 accuracy) is a noise filter for journalists to discover breaking news in Twitter (Liu et al., 2016): machine learning models for noise filtering and event detection are implemented, NLP is also used. "Fake News" Classifier (www.classify.news) allows to score the veracity of an article by entering its URL. The corpus articles are based on OpenSources labels. The tool focuses on NLP techniques and considers both content (bag of words; multinomial Naive Bayes classifier) and context (sentiment analysis, capitalization and punctuation usage; Adaptive Boosting). HeroX Fact Check Challenge (https://herox.com/factcheck/community) (2016-2017) and FakeNewsChallenge (http://www.fakenewschallenge.org/) (2017) competitions were held to help to create fact checking systems.

As to the winners of HeroX Fact Check Challenge, Fact Scan (1st place) can check several types of claims automatically, such as numerical, position, quote and object property claims. Claim Buster (2nd place) is also able to check simple statements; it can match claims, and it is based on knowledge bases. As regards FakeNewsChallenge, the teams focused on the headline-text body relationships. Talos Intelligence team (1st place) used the ensemble classifier (gradient-boosted decision trees and deep convolutional neural network). Word embeddings, based on Google News pretrained vectors, were used for the neural network. Such features are informative for decision trees: the number overlapping words between the headline and body text; similarities measured between the word count, bigrams and trigrams; similarities measured after TF-IDF weighting and singular value decomposition. UCL Machine Reading system (3rd place) is based on lexical and similarity features fed through a multi-layer perceptron with one hidden layer. Features for checking headline-text body consistency contain three elements: a bag-of-words term frequency vector of the headline; a bag-of-words term frequency vector of the body; the cosine similarity of TF-IDF vectors of the headline and the text body.

Fake news may be identified on different levels. Usually they are combined, from lexics and semantics to syntax. Most studies focus on lexics and semantics and some syntax principles; discourse and pragmatics have rarely been considered (Rubin et al., 2015b) due to their complexity.

On lexics level, stylistic features (part of speech (POS), length of words, subjectivity terms etc.) can be extracted that help to apart tabloid news (they are similar to fake news) with 0.77 accuracy (Lex et al., 2010). Numbers, imperatives, names of media persons can be extracted from news headlines (Clark, 2014); the numbers of these keywords can be used as features for classification with SVMs or Naive Bayes Classifier (Lary et al., 2010). Psycholinguistics lexicons, for instance LIWC (Pennebaker and Francis, 1999), can be used in performing binary text classifications for truthful vs deceptive texts (0.70 accuracy) (Mihalcea and Strapparava, 1999) — for example, methods can be based on frequency of affective words or action words. On syntax level, Probability Context Free Grammars can be used (0.85-0.91 accuracy) (Feng et al., 2012). On pragmatics level, pronouns with antecedents in text are more often used in fake news' headlines (Blom and Hansen, 2015). On discourse level, rhetorical structures are used (Rubin et al., 2015b): vector space modeling application predicts whether a report is truthful or deceptive (0.63 accuracy) for English. Corpus consists of seriously fabricated news stories. So rhetorical structures and discourse constituent parts and their coherence relations are possible deception detection markers in English news.

As to facts in the described event, in (Sauri and Pustejovsky, 2012) model is based on grammatical fact description structures in English and kindred languages. It is implemented in De Facto, a factuality profiler for eventualities based on lexical types and syntax constructions. The FactBank, annotated corpus in English, was also created. FactMinder, a fact checking and analysis assistant based on information extraction, can help to find relevant information (Goasdoué et al., 2013). Knowledge networks like Wikipedia can be used for simple fact checking questions (Ciampaglia et al., 2015).

In (Hardalov et al., 2016) the combined approach for automatically distinguishing credible from fake news, based on different features and combining different levels, is presented: there are linguistic (n-gram), credibility-related (capitalization, punctuation, pronoun use, sentiment polarity), and semantic (embeddings and DBPedia data) features. It can be applied to Bulgarian. The accuracy is from 0.75 to 0.99 on 3 different datasets.

The impact of different features on deception detection was studied in recent works (Fitzpatrick et al., 2015; Rosso et al., 2017).

There are no automated deception detection tools for news reports for Russian, although the field of deception detection in written texts is studied on the Russian Deception Bank (226 texts). The majority of research parameters are related to POS tags, lexical-semantic group, and other frequencies of LIWC lexicon words. The classifier's accuracy is 0.68 (Litvinova et al., 2017). Hence, we should base the research for Russian on the experience of methods for other languages, keeping in mind linguistics, social and cultural circumstances.

3 Research Objective

The aim is to reveal differences between fake and truthful news reports using markers from different linguistics levels. We use POS tags, length of words, sentiment terms, punctuation on the lexics level. Deception detection requires understanding of complex text structures, so we use Rhetorical Structures Theory (RST) relations as markers on the discourse level. In two experiments we shall classify the texts from the definite corpus.

4 Data Collection Principles

There are no sources that contain verified samples of fake and truthful news for Russian, although the problem of fake news is annually discussed on conference "Media Literacy, Media Ecology, Media Education: Digital Media for the Future" (Moscow). There are no Factbanks, unbiased fact checking websites, crowdsourcing projects, lists of truthful/deceptive sources. We can rely only on the presented facts, on the factuality.

The daily manual monitoring of news lasted 24 months (June 2015-June 2017). Online newspapers in Russian were used as sources. For balance, texts were from diverse sources: well-known news agencies’ websites, local or topic-based news portals, online newspapers from different countries. News source mention was not included in text annotations to avoid biases. Blogs and social media texts, analytic journalism stories based on opinions (not on facts) were not taken. We selected only serious fabrications. News stories were carefully analyzed in retrospect when the factuality was already known, to avoid biased evaluation. In case of mutual contradictions in the reports about the same event, a report was added to fake cases if at the same time period in online media existed reports with unproven facts and with their truthful refutation. So it was an intended fake and not a journalist's mistake caused by lack of facts.

5 Corpus Details and Data Analysis

The corpus consists of news reports about 48 different topics, with equal number of truthful and deceptive texts to each topic (not more than 12 texts for one topic). It contains 174 texts. The whole number of tokens is 33049. The mean length of texts is 189.04 tokens, the median length is 168.5 tokens. The whole number of rhetorical relations in corpus is 3147. Mean number of rhetorical relations in text is 18.09, the median number is 16.5.

The corpus size is conventional for the initial research on the field of automated deception detection, especially if we use the discourse level of language analysis, because it still requires manual annotation. Discourse parsers exist most notably for English (RASTA, SPADE, HILDA, CODRA etc.), and researchers do not use them even for English corpora when they need precise results. For comparison, the dataset in the paper which describes automated deception detection for news reports, based on RST, includes 144 news reports that were tagged manually (Rubin et al., 2015b). Corpus in the research about the impact of discourse markers on argument units classification (Eckle-Kohler et al., 2015) consists of 88 documents, predominantly news texts.

We used the following 18 normalized lexical markers for each text: average length of tokens; type-token ratio; frequency of adverbs; frequency of adjectives; frequency of pronouns-adverbs; frequency of numerals-adjectives; frequency of pronouns-adjectives; frequency of conjunctions; frequency of interjections; frequency of numerals; frequency of particles; frequency of nouns; frequency of pronouns-nouns; frequency of verbs; frequency of all punctuation marks; frequency of quotations; frequency of exclamation marks; frequency of lemmas from a sentiment lexicon.

All POS tags were obtained with MyStem tool for Russian which is for free use (some form words were excluded from the analysis). We collected seriously fabricated news reports, so we do not take capitalization as a feature. As there are no tools for sentiment polarity for Russian for free use, we use frequencies of lemmas from a list of

5000 sentiment words from reviews (Chetviorkin and Loukachevitch, 2012).

As to the discourse part, RST framework (Mann and Thompson, 1988) represents text as an hierarchical tree. Some parts are more essential (nucleus) than others (satellite). Text segments are connected to each other with relations. The theory pretends to be universal for all languages, so we chose it for our research. There are no discourse parsers for Russian: tagging and validation were made manually. We used UAM CorpusTool for discourse-level annotation. We based the research on the "classic" set by Mann and Thompson and added to it some more types: so, we created 4 types of Evidence according to the precision of source of information mention. News reports usually have a definite template, so a rather small number of relations was used. We have 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'. To avoid subjectivity of annotators' interpretation, we had 2 annotators and tried to solve this problem by preparing a precise manual for tagging and by developing consensus-building procedures. We selected Krippendorff's unitized alpha (0.78) as a measure of inter-annotator agreement.

The first dataset is based on statistics data about frequencies of lexical markers for each news report. The second one is based on statistics data about types of RST relations and their frequencies for each news report. In fact, we have a 'bag of relation types', disregarding their order.

We selected two supervised learning methods for texts classification and machine learning: Support Vector Machines (SVMs) and Random Forest, both realized in scikit-learn library for Python. SVMs were used with linear kernel and with rbf kernel. In both experiments (for both datasets) we used 10-fold cross-validation for estimator performance evaluation.

The baseline for all experiments is 50%, because there is the equal number of truthful and deceptive texts in the corpus.

6 Statistical Procedures

The results of two experiments are presented in Table 1.

	Precision	**Accuracy**	**Recall**	**F-measure**
Support Vector Machines, rbf kernel, 10-fold cross-validation				
Lexical features	**0.62**	**0.64**	**0.73**	**0.65**
Discourse features	0.56	0.54	0.52	0.51
Support Vector Machines, linear kernel, 10-fold cross-validation				
Lexical features	0.62	0.61	0.62	0.60
Discourse features	0.54	0.53	0.51	0.50
Random Forest Classifier, 10-fold cross-validation				
Lexical features	0.58	0.56	0.47	0.50
Discourse features	0.62	0.57	0.52	0.54

Table 1: Results for lexical and discourse features

We can evaluate that for the first one the classification task is solved better by SVMs (rbf kernel). The most significant features are: average length of tokens, frequency of sentiment words, frequency of particles, frequency of verbs. It was checked with Student's t-test. Although the results of the first experiment are better, for the second one the classification task is solved better by Random Forest Classifier. The most significant rhetorical relation types among discourse features are disjunction/conjunction, non-volitional cause, evaluation, elaboration. Non-volitional cause, elaboration, evaluation, conjunction are more typical for deceptive texts. Probably authors of fake news pay more attention to the causation, because they want to explain an event with the internal logic, without any inconsistencies.

7 Discussion

Automated deception detection seems to be a promising and methodologically challenging research topic, and further measures should be taken to find features for deception/truth detection in automated news verification model for Russian.

The model should be developed, learned and tested on larger data collections with different topics. We should use a complex approach and combine lexics and discourse methods, also combining them with other linguistics and statistical methods. For instance, n-grams, word embeddings, psycholinguistics features; syntactic level features on top of sequences of discourse relations should be studied. 'The trees' - hierarchies of RST relation types in texts should also be considered, to get better results. The extrapolation of the existing model to all possible news reports in Russian would be incorrect. But it can already be used as a preliminary filter for fake news detection. Results of its work should be double-checked, especially for suspicious instances. The model is also restricted by the absence of tools and corpora for Russian, as typical for NLP tasks for Russian. The guidelines for gathering a corpus of obviously truthful/deceptive news should also be improved.

8 Conclusions

News verification and automated fact checking tend to be very important issues in our world, with its information warfare. The research is initial. We collected a corpus for Russian (174 news reports, truthful and fake). We held two experiments, for both we applied SVMs algorithm (linear/rbf kernel) and Random Forest to classify the news reports into 2 classes: truthful/deceptive. We used 18 markers on lexics level, mostly frequencies of POS tags in texts. On discourse level we used frequencies of RST relations in texts. The classification task in the first experiment is solved better by SVMs (rbf kernel) (f-measure 0.65). The model based on RST features shows best results with Random Forest Classifier (f-measure 0.54) and should be modified. In the next research, the combination of different deception detection markers for Russian should be taken in order to make a better predictive model.

References

H. Allcott and M. Gentzkow. 2017. Social Media and Fake News in the 2016 Election. In *Journal of Economic Perspectives*. Vol. 31-2: 211-236.

J.N. Blom and K.R. Hansen. 2015. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76: 87-100.

I.I. Chetviorkin and N.Y. Loukachevitch. 2012. Extraction of Russian Sentiment Lexicon for Product Meta-Domain. In *Proceedings of COLING 2012: Technical Papers*: 593–610.

GL Ciampaglia, P. Shiralkar, LM Rocha, J. Bollen, F. Menczer, and A. Flammini. 2015. *Computational Fact Checking from Knowledge Networks*. PLoS ONE 10(6): e0128193. https://doi.org/10.1371/journal.pone.0128193

R. Clark. 2014. *Top 8 Secrets of How to Write an Upworthy Headline*, Poynter, URL: http://www.poynter.org/news/media-innovation/255886/top-8-secrets-of-how-to-write-an-upworthy-headline/

J. Eckle-Kohler, R. Kluge, I. Gurevych. 2015. On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse, In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP):* 2236-2242.

S. Feng, R. Banerjee, and Y. Choi. 2012. Syntactic Stylometry for Deception Detection. In *Proceedings 50th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics*, Vol. 2: Short Papers: 171–175.

E. Fitzpatrick, J. Bachenko, and T. Fornaciari. 2015. *Automatic Detection of Verbal Deception. Synthesis Lectures on Human Language Technologies.* Morgan & Claypool Publishers.

F. Goasdoué, K. Karanasos, Y. Katsis, J. Leblay, I. Manolescu, and S. Zampetakis. 2013. Fact Checking and Analyzing the Web. In *SIGMOD - ACM International Conference on Management of Data*, Jun 2013, New York, United States.

M. Hardalov, I. Koychev, P. Nakov. 2016. In Search of Credible News. In *Artificial Intelligence: Methodology, Systems, and Applications*: 172-180.

D.J. Lary, A. Nikitkov, and D. Stone. 2010. *Which Machine-Learning Models Best Predict Online Auction Seller Deception Risk?* American Accounting Association AAA Strategic and Emerging Technologies.

E. Lex, A. Juffinger, and M. Granitzer. 2010. Objectivity classification in online media. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*: 293-294.

O. Litvinova, T. Litvinova, P. Seredin, Y. Lyell. 2017. Deception Detection in Russian Texts. In *Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics*: 43-52.

X. Liu, Q. Li, A. Nourbakhsh, R. Fang, M. Thomas, K. Anderson, R. Kociuba, M. Vedder, S. Pomerville, R. Wudali, R. Martin, J. Duprey, A. Vachher, W. Keenan, and S. Shah. 2016. Reuters Tracer: A

Large Scale System of Detecting & Verifying Real-Time News Events from Twitter. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. Indianapolis, Indiana, USA, October 24-28, 2016: 207-216.

W.C. Mann and S.A. Thompson. 1988. *Rhetorical Structure Theory: Toward a Functional Theory of Text Organization*, Text, vol. 8, no.3: 243-281.

R. Mihalcea and C. Strapparava. 1999. The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In *Proceedings 47th Annual Meeting of the Association for Computational Linguistics*, Singapore: 309-312.

J. Pennebaker and M. Francis. 1999. *Linguistic inquiry and word count: LIWC*, Erlbaum Publishers.

P. Rosso and L. Cagnina. 2017. Deception Detection and Opinion Spam. In: *A Practical Guide to Sentiment Analysis*, Cambria, E., Das, D., Bandyopadhyay, S., Feraco, A. (Eds.), Socio-Affective Computing, vol. 5, Springer-Verlag: 155-171.

V.L. Rubin, N.J. Conroy, and Y.C. Chen. 2015b. *Towards News Verification: Deception Detection Methods for News Discourse*. In Proceedings of the Hawaii International Conference on System Sciences (HICSS48) Symposium on Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium, January 5-8, 11 pages.

V.L. Rubin, N.J. Conroy, and Y. Chen. 2015b. *Deception Detection for News: Three Types of Fakes*. Conference: ASIS T2015, At St. Louis, MO, USA.

R. Sauri and J. Pustejovsky. 2012. Are You Sure That This Happened? Assessing the Factuality Degree of Events in Text. In *Computational Linguistics*: 1-39.