ГЛУБОКИЕ РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ ДЛЯ АСПЕКТНО-ОРИЕНТИРОВАННОГО АНАЛИЗА ТОНАЛЬНОСТИ ОТЗЫВОВ ПОЛЬЗОВАТЕЛЕЙ НА РАЗЛИЧНЫХ ЯЗЫКАХ

Тарасов Д. С. (dtarasov3@gmail.com)

Интернет-портал reviewdot.ru, Казань, Россия

Ключевые слова: рекуррентные нейронные сети, анализ тональности, извлечение аспектных терминов, унифицированный подход

DEEP RECURRENT NEURAL NETWORKS FOR MULTIPLE LANGUAGE ASPECT-BASED SENTIMENT ANALYSIS OF USER REVIEWS

Tarasov D. S. (dtarasov3@gmail.com)

Reviewdot research,Kazan,Russian Federation

Deep Recurrent Neural Networks (RNNs) are powerful sequence models applicable to modeling natural language. In this work we study applicability of different RNN architectures including uni- and bi-directional Elman and Long Short-Term Memory (LSTM) models to aspect-based sentiment analysis that includes aspect terms extraction and aspect term sentiment polarity prediction tasks. We show that single RNN architecture without manual feature-engineering can be trained to do all these subtasks on English and Russian datasets. For aspect-term extraction subtask our system outperforms strong Conditional Random Fields (CRF) baselines and obtains state-of-the-art performance on Russian dataset. For aspect terms polarity prediction our results are below top-performing systems but still good for many practical applications.

Keywords: recurrent neural networks, sentiment polarity, aspect term extraction, unified approach

1. Introduction

In many practical natural language processing (NLP) systems, it is desirable to have one architecture that can be quickly adapted to different tasks and languages without the need to design new feature sets. Recent success of deep neural networks in general and deep RNNs in particular offers hope that this goal is now within reach. RNNs were applied to a number of English NLP problems, demonstrating their superior capabilities in slot-filling task [Mesnil et al, 2013] and opinion mining [Irsoy and Cardie, 2014].

While these results are promising it is still unclear if RNNs can now be used to replace other models in practical multi-purpose NLP system and if single RNN architecture can efficiently perform many different tasks.

Our work evaluates a number of RNN architectures on three different datasets: ABSA Restaurants (English) dataset from SemEval-2014 [Pontiki et al, 2014] and two Russian datasets (Restaurants and Cars) from SentiRuEval-2015.

We show that RNN performance on aspect terms extraction is close to state-of-the art and results on sentiment prediction, while being significantly behind top performing systems, outperform strong baselines and offer sufficient performance for use in practical applications. We discuss factors that contribute to RNNs results and suggest possible directions to further improve their performance on these tasks.

2. Related work

Sentiment analysis or opinion mining is the computational study of people's attitudes toward entities. In user reviews analysis two principal tasks are aspect terms extraction and aspect sentiment polarity prediction.

Aspect term extraction methods could roughly be divided into supervised and unsupervised approaches. In supervised approach aspect extraction is usually seen as sequence labeling problem, and often solved using variants of conditional random field (CRF) [Ganu et al, 2009;Breck and Cardie, 2007] methods, including semi-CRF systems, that operate at the phrase level and thus allow incorporation of phrase-level features [Choi and Cardie, 2010]. Such systems currently hold state-of-the arts results in term extraction from user reviews [Pontiki et al, 2014]. However, success of CRF and semi-CRF approaches depends on the access to rich feature sets such as dependency parse trees, named-entity taggers and other preprocessing components, that are often not readily available in underresourced languages such as Russian. Unsupervised approaches to term extraction attempts to cut cost and effort associated with manual feature selection and annotation of training data. These approaches typically utilize topic models such as Latent Dirichlet Allocation to learn aspect terms [Brody and Elhadad, 2010]. Their performance however, is below that of supervised systems trained on in-domain data.

Quite recently recurrent neural network models were proposed to solve sequence tagging problems, including similar opinion mining task [Irsoy and Cardie, 2014], demonstrating results superior to all previous systems. Importantly, these results were obtained using only word vectors as features, eliminating the need for complex feature-engineering schemes.

Similarly, sentiment polarity prediction subtask is solved within supervised and unsupervised learning frameworks. State-of-the-art performance on term polarity detection is currently obtained by using support vector machines (SVM) with rich feature sets that include parse trees and large opinion lexicons, together with preprocessing to resolve negation [Pontiki et al, 2014]. Unsupervised methods in sentiment analysis usually focus on construction of polarity lexicons for which number of approaches currently exists [Brody and Elhadad, 2010], and then applying heuristics to determine term polarity.

Neural network based methods were developed recently to detect document level and phrase-level sentiment, including tree-based autoencoders [Socher et al, 2011;2013] and convolutional neural networks [dos Santos and Gatti, 2014;Blunsom et al, 2014] and Elman-type RNNs were applied to sentence-level sentiment analysis with promising results [Wenge et al, 2014].

3. Methodology

3.1. Datasets

SemEval-2014 ABSA Restaurants dataset [Pontiki et al, 2014] was downloaded through MetaShare (http://metashare.ilsp.gr:8080/). This dataset is a subset of (Ganu et al, 2009) dataset. It contains English statements from restaurants reviews (3041 in training and 800 sentences in test set) annotated for aspect terms occurring in the sentences, aspect term polarities, and aspect category polarities.

Russian Restaurants dataset and corresponding Cars dataset released by SentiRuEval-2015 organizers to participants consist of similarly annotated reviews in Russian with a number of important differences. These datasets contain whole reviews, rather than individual sentences and are annotated with three categories of aspect terms “explicit” (roughly equivalent to SemEval-2014 notion of aspect term), “implicit” and so called “polarity facts”—statements that don't contain explicit judgments but nevertheless tell something good or bad about aspect in question.

Auxiliary dataset for training Russian unsupervised word vectors was constructed from concatenation of unannotated cars and restaurants reviews, provided by SentiRuEval-2015 organizers and 300,000 user reviews of various consumer products from reviewdot.ru database (obtained by crawling more than 200 online shops and catalogs).

3.2. Evaluation of human disagreement

As a part of this work we decided to evaluate human disagreement on SentiRuEval-2015 Restaurants dataset because we found many examples that seemed ambiguous. To do this we split dataset in two parts (70/30) and appointed two human judges. Human judges were given “annotation guidlines” sent by SentiRuEval organizers and 70% of annotated dataset. They then were asked to annotate remaining 30% with aspect terms (explicit, implicit and polar facts) and results were compared to original annotation using evaluation metrics described in “metrics” section.

3.3. Recurrent neural networks

A recurrent neural network [Elman, 1990] is a type of neural network that has recurrent connections. This makes them applicable for sequential prediction tasks, including NLP tasks. In this work, we consider simple Elman-type networks and Long-Short Term Memory architectures.

3.3.1. Simple recurrent neural network

In an Elman-type network (Fig. 1a), the hidden layer activations $h(t)$ at time step t are computed by transformation of the current input layer $x(t)$ and the previous hidden layer $h(t-1)$. Output $y(t)$ is computed from the hidden layer $h(t)$.

More formally, given a sequence of vectors {x (t)} where t = 1..T, an Elman-type RNN computes memory and output sequences:

$$h(t) = f(Wx(t) + Vh(t-1) + b) \tag{1}$$

$$y(t) = g(Uh(t) + c) \tag{2}$$

where f is a nonlinear function, such as the sigmoid or hyperbolic tangent function and g is the output function. W and V are weight matrices between the input and hidden layer, and between the hidden units. U is the output weight matrix, b and c are bias vectors connected to hidden and output units. $h(0)$ in equation (1) can be set to constant value that is chosen arbitrary or trained by backpropagation.

Deep RNN can be defined in many possible ways [Pascanu et al, 2013], but for the purposes of this work deep RNNs were obtained by stacking multiple recurrent layers on top of each other.

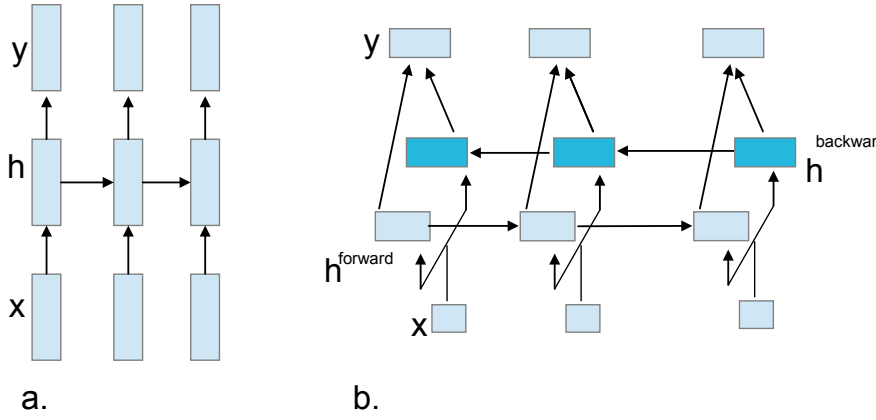

Figure 1. Recurrent neural networks, unfolded in time in three steps
a. Simple recurrent neural network
b. Bidirectional recurrent neural network

3.3.2. Long Short Term Memory

The structure of the LSTM [Hochreiter and Schmidhuber, 1997] allows it to train on problems with long term dependencies. In LSTM simple activation function f from above is replaced with composite LSTM activation function. Each LSTM hidden unit is augmented with a state variable $s(t)$ The hidden layer activations correspond to the 'memory cells' scaled by the activations of the 'output gates' o and computed in following way:

$$h(t) = o(t) * f(c(t)) \tag{3}$$

$$c(t) = d(t) * (c(t-1) + i(t)) * f(Wx(t) + Vh(t-1) + b) \tag{4}$$

where * denotes element-wise multiplication, $d(t)$ is dynamic activation function that scales state by "forget gate" and $i(t)$ is activation of input gate.

3.3.3. Bidirectional RNNs

In contrast with regular RNN that can only consider information from past states, bidirectional recurrent neural network (BRNN) [Schuster and Kuldip, 1997] can be trained using all available input data in the past and future. In BRNN (Fig. 1b) neuron states are split in a part responsible for positive time direction (forward states) and and a part for the negative time direction (backward states):

$$h(t)^{forward} = f(W^{forward} x(t) + V^{forward} h^{forward}(t-1) + b^{forward}) \tag{5}$$

$$h(t)^{backward} = f(W^{backward} x(t) + V^{backward} h^{backward}(t+1) + b^{backward}) \tag{6}$$

$$y(t) = g(U^{forward} h^{forward} + U^{backward} h^{backward} + c) \tag{7}$$

3.3.4. Training

All networks were trained using backpropagation through time (BPTT) [Werbos, 1990] algorithm with mini-batch gradient descent with one sentence per mini-batch as suggested in [Mesnil et al, 2013]. For sequence labeling tasks loss function was evaluated at every timestep, while for classification tasks such as term polarity prediction, loss function was only evaluated at the position corresponding to terms whose polarity was being predicted.

3.3.5. Regularization

To prevent overfitting small Gaussian noise was added to network inputs. Large networks were also regularized with dropout [Hinton et al, 2012] a recently proposed technique that omits certain proportion of the hidden units for each training sample.

3.4. Word embeddings

Real-valued embedding vectors for words were obtained by unsupervised training of Recurrent Neural Network Language Model (RNNLM) [Mikolov et al, 2010]. English embeddings of size 80 trained on 400M Google News dataset were downloaded

from RNNToolkit (http://rnnlm.org/) website. Russian embeddings of same size were trained using auxiliary dataset described above, using same method. Russian text was preprocessed by replacing all numbers with #number token and all occurrences of rare words were replaced by corresponding word shapes.

3.5. Evaluation metrics

For term extraction tasks where term boundaries are hard to identify even for humans, it is generally recommended to use soft measures like Binary Overlap that counts every overlapping match between a predicted and true expression as correct [Breck et al, 2007], and Proportional Overlap that computes partial correctness proportional to the overlapping amount of each match [Johansson and Moschitti, 2010].

From the description of SemEval-2014 task it appears that exact version of F- measure was used (only exact matches count), even though organizers note that "In several cases, the annotators disagreed on the exact boundaries of multi-word aspect terms".

For Russian SentiRuEval-2015 datasets, due to somewhat different annotation approach, multi-word (4 and 5 word terms) are quite common and human disagreement is quite large (as will be shown below). SentiRuEval-2015 organizers adopt two metrics for aspect-term extraction—main (based on exact count) and secondary (based on proportional overlap).

In SentiRuEval-2015 datasets all terms are tagged as "relevant" (related to target entity), or irrelevant (related to something else) and official metrics only count identification of relevant terms as correct. We feel that identification of aspect term and classification it as "relevant" or not are two fundamentally different tasks and should be measured separately. Due to extremely low presence (less than 5%) of irrelevant terms, their exclusion is quite hard for machine learning algorithm to achieve, and finding algorithms that do that well is a problem of significant theoretical interest. Such systems cannot be identified using official metrics, since contribution of "relevance" detection to overall F1 value is rather small.

For the purposes of this paper unless otherwise stated, we apply F-measure based on proportional overlap to facilitate comparison of results obtained on different datasets. For English Restaurants ABSA dataset F-measure is computed on Test dataset of 800 sentences (that was not used in development of models). For Russian datasets, as test data were not available at the time of this work, we separate development set of 5000 words and use 7-fold cross-validation on remaining data, similar to [Isroy and Cardie, 2014] approach. Since we participated in a number of SentiRuEval-2015 tracks, official results according to SentiRuEval-2015 metrics are also shown for comparison and discussion purposes.

For classification tasks such as sentiment polarity and aspect category detection tasks, macro average of F-measure cannot be used due to the fact that some categories (such as "conflict" polarity, named "both" in Russian dataset) are extremely rare (Russian Restaurant dataset contains less than 80 instances of "both" polarity per 3000 instances of aspect terms). F-measure for such categories is subject to huge sampling error, and can also be undefined (with zero precession and recall), making macro

average value undefined also. To prevent this problem from occurring SemEval-2014 uses Accuracy instead of F-measure. SentiRuEval-2015 organizers use F1 micro average in addition to macro average. In this paper, for classification tasks we show overall accuracy, computing macro-average as additional measure where possible.

3.6. Baselines

For term extraction task we consider several baseline systems: simple feed-forward multi-layer perceptron (MLP), frame-level MLP (a feed-forward MLP with inputs of only word embedding features within a word context window), logistic regression using word embedding features, and CRF using stemmed words and POS-tags as features.

4. Results and Discussion

4.1. Aspect term extraction task

Tables 1–3 summarize our results on aspect term extraction. Initially, for Russian Restaurant dataset, we found it very difficult to improve upon simple CRF baseline. Manual examination of annotation revealed a number of inconsistent decisions in provided training data, for example in one place term “официантка Любовь” (“servant Lubov”) was tagged as a whole, while in other similar case servant name was not tagged as part of the term. That led us to evaluation of human disagreement that appeared to be very close to baseline results, making term extraction very formidable challenge.

Nevertheless, we found that augmented forward RNN outperforms CRF baseline on explicit aspect extraction and deep LSTM model outperforms both CRF and Frame-NN baselines on all subtasks, while simple BRNN while providing reasonable good results, failed to improve on these baselines in contrast with English dataset. We think that inconsistent annotation in training set leads to over-fitting in simple BRNNs, because complex local models are learned before long time dependencies in the data can be discovered.

Overall, as shown in Table 2, our system obtains best result in extraction of all aspects terms according to proportional measure and best result in extraction of all aspect terms on cars dataset according to exact measure, while holding second-best result on restaurants dataset. These good results, should, however, be interpreted with caution due to relatively small number of participants, general lack of strong competitors and poor quality of the data (at least in Restaurant domain).

Therefore, to better understand system capabilities we evaluated our system on English dataset of SemEval-2014. The advantage of this dataset is that it is carefully cleaned from errors and also results of state-of-the-art systems are readily available for comparison. Table 3 demonstrates that in this dataset our system did not obtain top results. Still, LSTM performance is quite good (equivalent to 6th best result of 28 total participants).

Table 1. F-measure (proportional overlap) on SentiRuEval dataset, evaluated using 7-fold cross-validation

Mehod	SentiRuEval Restaurants dataset				SentiRuEval Cars dataset			
	Explicit	Implict	Fact	Macro average	Explicit	Implict	Fact	Macro average
Human Judge 1	69.1	58.7	33.0	53.6	—	—	—	—
Human Judge 2	65.0	62.3	27.0	51.4	—	—	—	—
CRF baseline	68.2	57.7	24.0	49.96	—	—	—	—
Logistic regression	54.0	43.0	3.0	33.3	70.1	75.4	15.2	53.6
MLP	64.5	53.6	18.2	45.3	75.8	82.2	34.8	64.2
Frame-NN	67.9	61.4	**26.1**	51.8	76.0	**83.0**	33.0	64.0
Simple RNN	68.4	58.5	20.0	48.9	75.2	81.3	30.1	62.2
Simple RNN augmented with one future word	68.9	60.0	25.3	51.4	75.8	82.0	31.4	63.1
Simple RNN augmented with one future word + dropout	71.1	56.0	20.1	49.06	76.0	82.1	24.3	60.8
Bidirectional RNN	69.8	61.2	19.1	50.3	76.1	81.5	32.1	63.2
Bidirectional LSTM	**73.5**	**64.3**	23.5	**53.76**	**77.0**	82.5	**36.3**	**65.3**

Table 2. F-measure on SentiRuEval Test dataset (according to SentiRuEval results)

Method	SentiRuEval Restaurants dataset				SentiRuEval Cars dataset			
	Proportional		Exact		Proportional		Exact	
	Explicit	All	Explicit	All	Explicit	All	Explicit	All
BRNN	67.2	52.2	57.5	64.5	71.7	70.4	61.7	59.9
LSTM	71.9	**60.0**	62.6	**66.8**	—	—	—	—
LSTM, Depth 2	—	—	—	—	**74.8**	**71.4**	65.1	63.0
Other systems best result	**72.8**	59.6	**63.1**	59.5	73.0	65.9	**67.6**	**63.6**

Table 3. Results on English SemEval ABSA Restaurant dataset (computed by us, using SemEval official metrics), reference results are taken from [Pontiki et al, 2014]

Method	F1 value
baseline	47.15
CRF with words and POS tags features	75.20
6th-best result	79.60
Top result	84.01
BRNN	76.20
LSTM	79.80

4.2. Sentiment polarity prediction task

Tables 4–6 summarize sentiment polarity results. Here more complex systems generally obtain superior results to simpler methodologies.

Using SentiRuEval-2015 official metrics we obtain second-best result in explicit aspect term polarity prediction on cars-dataset and third-result in restaurants dataset (unfortunately, results from our top systems were not included in official results due to errors that we made in data format. This error only became apparent after release of test sets and thus impossible to correct). Also, relatively poor results are partially explained by the fact that our system was optimized to all-term polarity prediction task, leading to suboptimal performance on explicit-term only task (information about official metrics were released by organizers with delay and we were not able to adapt all systems due to time and resource constraints). On English ABSA Restaurant dataset we obtain accuracy of 69.7, significantly below best results, but still reasonable.

Even through our results here are below top systems, they are reasonable good and have some theoretical value in demonstrating that exactly same architecture can be used both for sequence tagging and polarity prediction tasks. It also worth noting, that we used neither sentiment lexicon, nor special preprocessing steps for negation (we found that RNNs under certain conditions are capable to learn negation just from training data). Another important finding here that using hidden layer activations of RNNLM model as features instead of word vectors considerably improves overall system performance. Our hypothesis is that next-word prediction task of RNNLM includes the need to understand word dependencies—a knowledge that shown to be crucial in aspect-term polarity prediction task. This knowledge from unsupervised model can thus be leveraged by supervised RNN to enhance performance.

Table 4. Results on all-terms polarity prediction task on SentiRuEval dataset (F1 macro average on positive and negative classes and overall accuracy over all terms)

	Restaurants		Cars	
Method	**Macro F1**	**Accuracy**	**Macro F1**	**Accuracy**
TDNN N=3	61.0	57.4	55.2	56.2
RNN	63.1	59.2	57.1	57.1
BRNN	67.4	60.3	60.3	56.9
LSTM	70.2	61.1	62.4	58.0
LSTM + RNNLM features *	74.1	62.5	65.0	59.1

* Obtaining by using hidden layer activations of RNNLM

Table 5. Results on explicit-only terms polarity classification (according to SentiRuEva-2015l official results)

Method	**Restaurants**	**Cars**
BRNN	61.9	64.7
LSTM + RNNLM features	—	65.3
Top result	82.4	74.2

Table 6. Results for English terms polarity classification on ABSA Restaurants SemEval-2014 dataset (according to our evaluation metrics)

Method	Accuracy
Baseline	64.00
Sentiment lexica over dependency graphs *	69.50
BRNN	65.10
LSTM	69.70
Top result	82.92

* Value taken from [Wettendorf et al, 2015]

5. Conclusions

In aspect term extraction task recurrent neural networks models demonstrate excellent perfomance. On Russian SentiRuEval-2015 dataset our system obtained best result in extraction of all aspects terms according to proportional measure and best result in extraction of all aspect terms on cars dataset according to exact measure, while holding second-best result on restaurants dataset. On English SentEval-2014 dataset, we obtained reasonable good results, equivalent to 6th best known result on this dataset. From all RNN models, best results were obtained with deep bidirectional LSTM with 2 hidden layers.

For aspect term polarity predictions, we obtained second best result on SentiRuEval-2015 car dataset and third best result on SentiRuEval-2015 car restaurants dataset. We also obtained good results on all terms polarity prediction. To our knowledge, this is first time when LSTM models were applied to aspect term polarity prediction with reasonable good results.

Overall, our work demonstrates that RNN models are useful in aspect-based sentiment analysis and can be utilized for rapid prototyping and deployment of opinion mining systems in different languages.

Acknowledgments

Author want to thank Ekatertina Izotova for help with data format conversion, anonymous reviewers for helpful comments and SentiRuEval organizers for preparing and running evaluation and thus making this work possible.

References

1. *Blunsom, P., Grefenstette, E., & Kalchbrenner, N.* (2014). A convolutional neural network for modelling sentences. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics.
2. *Breck E., Choi Y., Cardie C.* (2007). Identifying expressions of opinion in context. In IJCAI, pp. 2683–2688.
3. *Brody S., Elhadad N.* (2010). An unsupervised aspect-sentiment model for online reviews. In Proceedings of NAACL, pp. 804–812, Los Angeles, California
4. *Choi Y., Cardie C.* (2010). Hierarchical sequential learning for extracting opinions and their attributes. In Proceedings of the ACL 2010 Conference Short Papers, pp. 269–274.
5. *dos Santos, C. N., & Gatti, M.* (2014). Deep convolutional neural networks for sentiment analysis of short texts. In Proceedings of the 25th International Conference on Computational Linguistics (COLING), Dublin, Ireland.
6. *Elman J.* (1990). Finding structure in time. Cognitive science, 14(2):179–211.
7. *Ganu, G., Elhadad, N., & Marian, A.* (2009, June). Beyond the Stars: Improving Rating Predictions using Review Text Content. In WebDB (Vol. 9, pp. 1–6).
8. *Hinton G. E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R.* (2012). Improving neural networks by preventing coadaptation of feature detectors. arXiv preprint arXiv:1207.0580
9. *Hochreiter, S., & Schmidhuber, J.* (1997). Long short-term memory. Neural computation, 9(8), 1735–1780.
10. *Irsoy O., Cardie C.* Opinion Mining with Deep Recurrent Neural Networks (2014). EMNLP, Doha, Qatar. pp. 720–728
11. *Johansson R., Moschitti A.* (2010). Syntactic and semantic structure for opinion expression detection. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning, pp. 67–76. Association for Computational Linguistics.
12. *Mesnil, G., He, X., Deng, L. & Bengio, Y.* (2013). Investigation of recurrent neural network architectures and learning methods for spoken language understanding. In INTERSPEECH pp. 3771–3775 : ISCA.
13. *Mikolov T., Karafi´at M., Burget L., Cernock´y J., Khudanpur S.* (2010). Recurrent neural network based language model. In INTERSPEECH, pp. 1045–1048.
14. *Pascanu, R., Gulcehre, C., Cho, K., & Bengio, Y.* (2013). How to construct deep recurrent neural networks. arXiv preprint arXiv:1312.6026.
15. *Pontiki M., Papageorgiou, H., Galanis, D., Androutsopoulos, I., Pavlopoulos, J., & Manandhar, S.* (2014). Semeval-2014 task 4: Aspect based sentiment analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 27–35).
16. *Schuster M., Kuldip K. P.* (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11):2673–2681.
17. *Socher, R., Pennington, J., Huang, E. H., Ng, A. Y., & Manning, C. D.* (2011, July). Semi-supervised recursive autoencoders for predicting sentiment distributions. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 151–161). Association for Computational Linguistics.

18. *Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C.* (2013, October). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the conference on empirical methods in natural language processing (EMNLP) (Vol. 1631, p. 1642).
19. *Wenge R., Baolin P., Yuanxin O., Chao Li, Zhang X.* (2004) Structural information aware deep semi-supervised recurrent neural network for sentiment analysis. Frontiers of Computer Science, pp. 1–14, http://dx.doi.org/10.1007/s11704-014-4085-7
20. *Werbos, P. J.* (1990). Backpropagation through time: what it does and how to do it. Proceedings of the IEEE, 78(10), 1550–1560.
21. *Wettendorf C., Jegan R., Korner A., Zerche J.* (2014) SNAP: A Multi-Stage XML-Pipeline for Aspect Based Sentiment Analysis In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp. 578–584