Determining Sentiment and Important Properties of Ukrainian Language User Reviews

Dmytro Babenko and Vsevolod Dyomkin

Ukrainian Catholic University, Lviv, Ukraine

d.babenko@ucu.edu.ua, vseloved@gmail.com

Abstract. The main idea of this project to develop a model, which can recognize emotions from user review text. Many websites contains review text with marks (stars or another representation) about some products or services. Sometimes user can write only text without a mark, in this case, such a model can help us to understand the estimation of this user. It can be a review of some special product, about booking a hotel or renting a car or something like that. Furthermore, another purpose of this project to detect which characteristics were influenced for such review. Here, the model should detect sentiments from text and classify for positive and negative. Moreover, it will understand which reasons were caused by these emotions and why the user wrote such texts. As this project mostly works with text processing, it is classical natural language processing task using deep learning. Probably, as a result, it would be some neural network, which was selected as the best from several experiments.

Keywords: natural language processing · deep learning· neural net · sentiment · review · long short-term memory · recursive neural tensor network

1 Introduction and Motivation

Many services would like to understand whether their customers are satisfied or not. By these reasons, they usually ask to leave a review about their services or goods. In many cases, when selecting a service or when choosing a product the amount of information available in user reviews can be or greater volume and more trustworthy than the official product description provided by vendor [4]. The reviews can help them to be better and can attract new customers. Usually, the number of reviews for specific service item (i.e hotel, restaurant) can be overwhelming.

The review reader can gain an understanding of overall satisfaction by looking at summary statistics such as the average rating or score but the details about what makes service great or not are hidden inside the body of the reviews. With such a large amount of data to process, it would be beneficial to automate the process and generate the summary of the reviews. Furthermore, many reviews can help services to know which points are interested by customers. To get this knowledge, they need to analyze data

In: Proceedings of the 1st Masters Symposium on Advances in Data Mining, Machine Learning, and Computer Vision (MS-AMLV 2019), Lviv, Ukraine, November 15-16, 2019, pp. 106–114

more deeply. The model, which is proposed, will help analyze such kind of data and extract important information from it.

The model, which will be trained during this work, should classify and cluster derivation sentiments for the key aspects of a reviewed item. From the huge set of review texts, the approach solution output short texts (probably sentences) from the reviews for which user has strong positive and negative opinions about.

Customer reviews contain several forms of salient information. First, there are the rating stars, which probably one of the most important. But rating, even when broken out into categories - on Airbnb, for example, category include accuracy, communication, cleanliness, location, check-in, and value has zero explanatory power. So, there is review text, which tells a story, and stories sell, so we need to know the aspects of product or service discussed, the wording used to describe them, and the sentiment expressed.

How the analyzing of reviews can help? To analyze customer reviews, some services like hotels, restaurants can understand what their potential customers need. It allows them to pay more attention to the things, which are interested in customers now and realized that some of their points are irritated by users. In this case, they will be more familiar with customer needs and can be oriented more for consumers. Moreover, it can simply to analyze the specific review text left by the client, which don't have a star, and the analyzer can acquire which characteristics are positive and negative was in this text.

As a common factor, such kind of analyzers was developed for the English language, but we do not have a similar solution for Ukrainian, so it can help to develop analyzing such kind of texts in Ukraine.

The paper is structured as follows. In section 2, we discuss related work in the area of sentiment detection and analysis for English and Ukrainian language. We also review several solutions. In section 3, we provide more description of problems and analyze the research hypothesis. In section 4 and 5, we show how we are going to solve this problem, which methods we are going to use, and present the envisioned approach to the final solution. Furthermore, in this section, we give information about research, development, and evaluation. In section 6, we provide a detailed plan of work with expected temporary results. At the end, section 7 contains a summary.

2 Review of Related Work

There are not so many works for the Ukrainian language in sentimental analysis. Moreover, there were not found related works for analyzing reviews in Ukrainian about product or services.

Nevertheless, there were made many works for collecting corpus and creating embeddings for the Ukrainian language. It was done by Lank-UK open community of people passionate about NLP and intelligent text processing [1]. Additionally, there were made details research in for tone dictionary of the Ukrainian language [2].

However, it is possible to find several approaches to the English language. The idea can be similar for any other languages, but of course, there are some differences related to the corpus, embedding, sentimental keywords, and others. There are several startups, that analyze customer reviews [3]:

- Revezue focuses on products and product attributes in addition to product brands health, with a couple of differentiators. One is special attention to discovering different ways people talk about a given topic, and a second is the ability to identify sentiment in phrases that lack obvious clues like using words like “good,” “happy,” and “terrible.”
- Aspectiva provides aspect-based review aggregation and product search focused on reviewer’s perceptions of product attributes and capabilities. Aspect extraction, implemented via unsupervised machine learning, is coupled with behind-the-scenes analytics to reveal “the true uses of any product and generates recommendations based on the full user experience.”
- SmartMunk analysis centers on satisfaction drivers rather than on product and service attributes. The principle is that customer satisfaction drives business outcomes, so it is important to focus on elements that make customers happy or that disappoint.
- SentiGeek is a pre-launch customer feedback/review-sentiment company. Candidate markets – indicating the breadth of review-sentiment interest — include online retail, market research, marketing, financial institutions, and public administrations. The product is designed to support reporting, analysis, and monitoring options. It extracts opinion words and phrases, and fine-grained sentiment and provides analyses by review and by opinion holder, with the ability to generate customer profiles.

Unlike ready products, we are interested more about research work in this area:

- User Review Sentiment Classification and Aggregation [4]. This paper contains close research, which we are doing in this master thesis. There are described some methods which identify sentences that are representative of positive and negative aspects of the review subjects. In this paper, they propose an approach to review summarization that leverages trained models for classification and clustering to derive sentiment for the key aspects of a reviewed item. Given a set of item reviews, their system will output sentences from the reviews that are representative for the key aspects, regarding which users had strong positive or negative opinions about. There concluded that for classification they recommend a gradient boost tree over Naive Bayes classifiers. They also mentioned that their approach of utilizing the review score for sentence-level labels worked well for positive sentences, but was less effective for negative sentences. Furthermore, using POS tagging was unable to improve the performance of their clustering algorithm. Advanced NLP techniques may be of value to help improve the clustering component of this system.
- Deep learning for sentiment analysis of movie reviews [5]. In this paper, there is an exploration of various NLP methods for sentimental analysis. They consider two different datasets, one with binary labels, and one with multi-class labels. For the binary classification, they applied the bag of words, and skip-gram word2vec models followed by various classifiers, including random forest, SVM, and logistic regression. For the multi-class case, they implemented the recursive neural tensor networks (RNTN). For avoiding high computation cost, there is proposed low-rank RNTN and

provided some comparison results which can prove that low-rank RNTN gives no worse result than standard RNTN (see Fig. 1).

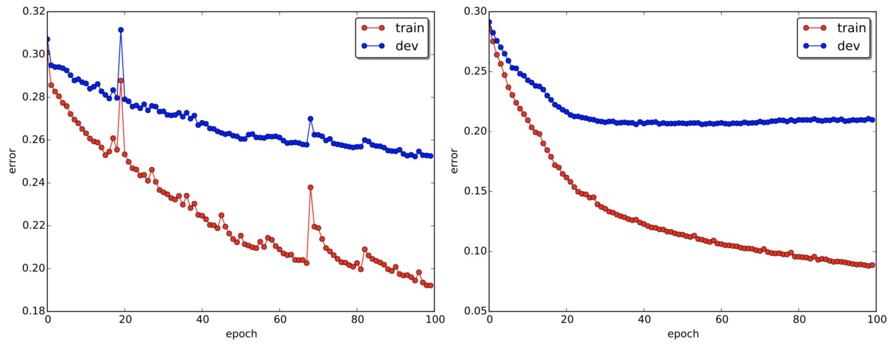

Fig.1. Evolution of the training and development errors for the standard RNTN (left) and the low-rank RNTN with rank = 5 (right)

There is also shown how they can take advantage of the fast and cheap convergence of the low-rank RNTN method by training multiple models and ensembling the predictions. For a given image, x and a model M_i , $P(x = y_j;M_i)$ represents the probability that the input image x is classified as $y_j(j = 1,2,...,m)$ under the model M_i. The simple ensemble average is then:

$$P(x = y_j; M_1, ..., M_n) = \frac{1}{n}\sum_{i=1}^{n} P(x = y_i; M_i)$$

– Generalized Autoregressive Pre-training for Language Understanding [6]. There is a paper with the capability of modeling bidirectional contexts, de-noising auto-encoding based pre-training, like BERT, which achieves better performance than pre-training approaches based on autoregressive language modeling. In light of these pros and cons, they propose XLNet, a generalized autoregressive pre-training method that enables learning bidirectional contexts by maximizing the expected likelihood over all permutations of the factorization order and overcomes the limitations of BERT thanks to its autoregressive formulation. There is a repository to track the progress in Natural Language Processing (NLP), including the datasets and the current state-of-the-art for the most common NLP tasks. It consists of models that are trained on Yelp Review Dataset for binary and fine-grained (five-class) classification.

We expect that, considering these papers, we can avoid some additional research works, which probably can give bad results. We also take a lot of significant advice on how it would be better and faster to train our models. Moreover, we use the ready corpus and embedding for the Ukrainian language, which simplifies this work.

3 Research Hypothesis and Problem

3.1 Problem

Initially, we have review text about some services (e.g., hotels, restaurants review). From this text, we need to understand the rank of this review. Usually, a single rank is not fully sufficiently informative, because we do not know what were the reasons caused to write such a review. As a result, we should understand which sentences have strong positive content and which are negative. As this problem is considered for services, the next steps, which should be performed, are done to understand what exactly the topic was in the specific sentence (for positive or negative meaning), its location, price, personnel or something like that.

Secondly, there are many reviews about hotels, restaurant, or something like that. We would like to have an instrument, which can analyze which characteristics are important nowadays for potential customers.

3.2 Example Solution

For instance, we have a review text about one hotel as imagine in Fig. 2. As we can see from the image above, the user set your rank, but the main information is hidden in the body of the text. In this case, we would like to know where the client was satisfied and was not. In Fig.3, there is separated positive (underline by the green line) and negative (red line) sentences. Furthermore, here should conclude that client was satisfied with the *room* of the hotel (the positive sentence is related to the room) and dissatisfied with the *food* (the negative sentence is related to the breakfast). As a result, this specific review can help understand services that they should improve some options for breakfast to be more attractive for clients.

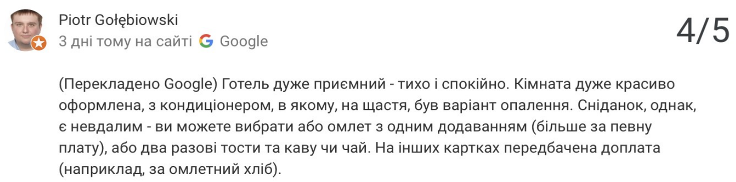

Fig. 2. Review example about one hotel

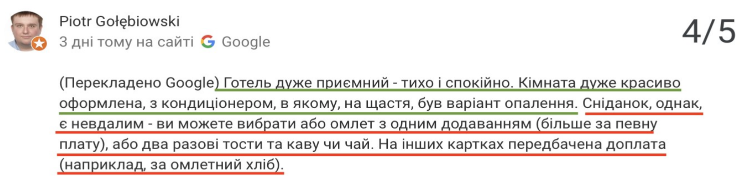

Fig. 3. Analyzed review example about one hotel

For about analyzing a lot of reviews such as in Fig. 2, let us imagine the situation that we have a lot of reviews about one (in this case it would be clear to know what exactly good or bad in this hotel) or about different hotels (it can help to understand what potential customers need). Correspondingly, we need general statistics about these review contents. For example, in the diagram below (Fig.4) we see how customers are satisfied with the following topics.

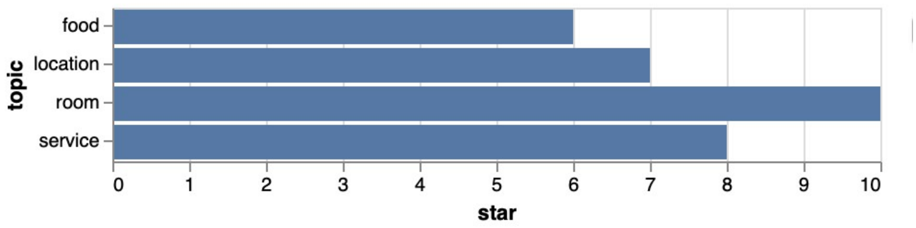

Fig. 4. Statistics about several reviews in each topics

As it was not mentioned in the first part, we would like to focus on getting a star for each topic. Hence, for one review, the model should detect the star for each specified topic.

To sum up, there should be one instrument, which can detect a star for one review, classify positive and negative text, and cluster topics (location, food, service, etc.) and detect star for each topic. Another instrument will analyze several reviews for one or different service item (hotel, restaurant, etc.).

4 Envisioned Approach to Problem Solution

4.1 Dataset

To create any model, we first need a relevant dataset for the Ukrainian language. For this work, we would like to focus on services review, initially for hotels, later consider restaurants. Probably, after that, we try to extend our model and use it for product reviews and something else. Hence, the most available reviews about hotels in Ukrainian are in *booking.com*, there are also enough *google* reviews. As training data reviews form *booking.com* can be used, and *google* reviews can be used to test. Of course, other resources should be considered to get such data.

In most cases, these reviews have star rank, but rarely, there is separated positive and negative text (it does in *booking.com* but does not for *google* reviews), so some dataset should be marked manually or create an instrument which can do it. Additional marking dataset is also needed for training the model – to detect what topic is important and what is the rank of this topic.

4.2 Baseline

Though baseline models do not give a good result, it always is better to have such a model. In the early stages, baseline models can help us to notice several issues, which can be noticed late in the more advanced model. Furthermore, it would be simpler to test some experiments quickly using the baseline instead of the advanced approach.

In most models, for this work, there would be baseline based on a bag of words and using simple classifier like logistic regression, random forest.

4.3 Advanced Models

As there are already corpus and embeddings for Ukrainian language, it opens a way for using the more advanced NLP models. In this research, it make sense to consider related work for English language, which was discussed in section 2. For classification (rank classification or classification into positive or negative texts), we have a plan to evaluate and compare several models like SVM, logistic regression and random forest. It also makes sense to try recursive neural tensor networks (RNTN) and long short-term memory (LSTM) using embeddings. Probably, LSTM would be the final model, we will see.

4.4 Evaluation

There also should be provided evaluation result that shows how the ready models work on specific data. Unlike the specific domain, it makes sense to provide statistics out of the domain (e.g., if the model was trained on hotels review data then try to use it on restaurant review data). In this part, we also would like to analyze the errors of models, where the model can be used in other domain.

4.5 Transfer Learning

When some model (it does not matter classifier for rank or detection important topics) is already done and works good for such kind of data, we are not sure that it would work on another data. In common case, some new data has appeared recently, but the model was trained. To avoid re-training, we can use transfer learning for adaption our model to new data. In this work, we are going to try fine-tuning model on new data (e.g., the model was trained on data from *booking.com* and we want to fine-tune for data from other resources). Furthermore, there should be provided comparison result of using the model on new data without fine-tuning and with it.

5 Plan

A tentative schedule for the research is given in Table 1.

Table 1. Tentative plan for related work

Milestone	**Start Date**	**End Date**
Review of past and current related work	Aug 2019	Aug 2019
Analyze data, scribe data	Aug 2019	Sep 2019
Thesis proposal	Sep 2019	Sep 2019
Baseline models	Sep 2019	Oct 2019
Advanced model	Oct 2019	Nov 2019
Evaluation model	Nov 2019	Nov 2019
Fine-tune existed model (transfer learning)	Nov 2019	Dec 2019
Writing thesis	Dec 2019	Jan 2020
Thesis defense	Jan 2020	Jan 2020

6 Early Result

Currently, there were analyzed some related works for the English language, which was described in section 2. In addition, there were implemented one crawler [7] for scribing several datasets from *booking.com* (see result on Fig. 5). Moreover, there is the initial implementation of the baseline model, which gave low accuracy, but it allowed us to understand some important things on early stages (e.g. usually positive reviews more than negative in the dataset, it influences for training process).

	title	pos_text	neg_text	ratingValue	bestRating
0	наступної мандрівки до Львова оберемо цей готель.	Відмінне розташування готелю! Дуже близько до ...	один маленький мінус, але нам це не завдало не...	10.0	10.0
1	Сніданок - великі порціїї, смачно!	Сніданок - великі порціїї, смачно! 1 страва бе...	Сам готель усередині нагадує гуртожиток чи хос...	6.3	10.0
2	Все чудово! За нагоди, обов'язково знову завіт...	Дуже чудовий і затишний отель! Дуже зручне роз...	Сантехніку (труби , стоки) необхідно перевірит...	9.6	10.0
3	класне розташування, такий собі острівець тиші...	класне розташування, такий собі острівець тиші...	начебто новий ремонт - а велика тріщина над дв...	7.5	10.0
4	Гарне місце для туристів або відрядження, але ...	Ідеальний номер, як для тризіркового готелю - ...	Дуже круті сходи. Складно знайти поруч місце д...	9.2	10.0
5	- Чудове розташування у тихій вуличці у самому...	- Чудове розташування у тихій вуличці у самому...	- Дуже круті сходи між поверхами і відсутність...	7.1	10.0

Fig.5. Dataset example

7 Summary

This proposal has defined the concept of detecting emotions and sentiments from user review text. Mostly, there is a focus on services reviews such a hotel or restaurants, but there is also mentioned trying to use this approach for a product review or any other domain.

The proposed work should be followed by the subjects, which are described here. It means to consider already related work, analyze data, create a baseline, try using different models with the defined approach, and make a conclusion.

References

1. Lang-UK. http://lang.org.ua/en/
2. Romanyuk, A., Romanyshyn, M.Y.: Tonalnyi slovnyk ukrains'koi movy na osnovi sentyment-anotovanogo korpusu. Ukrains'ke Movoznavstvo, **43**, 63–74 (2013) – in Ukrainian
3. AI startups that analyze customer review. https://venturebeat.com/2017/01/27/4ai-startups-that-analyze-customerreviews/
4. Garcia, S., Yin, P.: User Review Sentiment Classification and Aggregation.
5. Pouransari, H, Ghili, S.: Deep learning for sentiment analysis of movie reviews. https://cs224d.stanford.edu/reports/PouransariHadi.pdf
6. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., Le, Q.V.: XLNet: generalized autoregressive pretraining for language understanding. arXiv preprint arXiv:1906.08237 (2019)
7. GitHub repository. https://github.com/DmytroBabenko/Detectemotionsentimental