кітапханасының мысалында талдау жасалады, оның негізгі міндеті балалардағы қант диабетін анықтау болып табылады.

Кілт сөздер: expert, python, Knowledge Engine.

Lik E.V.
Scientific superviser: Kim E.R.

Expert systems as the field of intelligent information systems

Abstract. In this article the expert system is reviewed and its simplified architecture is given. The functioning of the expert system is analyzed with the example of «experta» library in thePython programming language, the main task of which is to detect diabetes in children.

Keywords: expert, python, Knowledge Engine

Сведения об авторах:
Лик Эдуард Валерьевич, студент второго курса университета «Туран».
Ким Екатерина Р., к.т.н., доцент кафедры ИТ университета «Туран».

УДК 530.1, 681.3.06

Zimanova D.A, Sadir A.K, Kassymov B.M
Kazakh -British technical university
Almaty, Kazakhstan
Scientific superviser: Shamoi P.S.

SENTIMENT ANALYSIS OF SOCIAL NETWORKING SERVICES DATA USING FUZZY SETS AND LOGIC

Abstract*. This paper is devoted to usage of fuzzy logic in sentiment analysis of social network comments. The main feature of our work is that we aim to provide not only novel algorithm for sentiment analysis, but also to approve the correctness of our method, we implemented a prototype application on iOS platform. Our target audience is bloggers, business owners, SMM managers, artists and even government workers.*

In today's world, huge value of information is created every day. However, all these data is unstructured, which means that, despite that Internet is a transparent platform for reading data, it is impossible to make sense out of big data. Using our application user will be able to access generalized public mood towards some topics from social network posts.

Key words*: Sentiment analysis, fuzzy logic, fuzzy sets, NLTK, text processing, Emoji Sentiment Ranking, emoji processing, fuzzification*

Nowadays text processing technologies have attracted more attention than ever before and social networks have a vast influence on human life. To this end, Sentiment Analysis is a rapidly developing field in Artificial Intelligence (AI), consisting of widely differing tools and evaluation methods which target fetching emotional tones and subjective point of view of a speaker or a writer in some piece of text. In other words, sentiment analysis refers to a set of technologies to measure how positive or negative the text is. In the current paper, on the one hand, we focus on introducing a novel algorithm for sentiment analysis which mainly relies on python's Natural Language Toolkit (NLTK), fuzzy mathematics and Emoji Sentiment Ranking. We propose the processing of emojis (picture-like symbols commonly used to express someone's emotions) to boost sentiment analysis of a certain text [1].

The decision to choose Instagram as the data source was made because it is one of the most spread social networks all around the world and the fact that it is a rich source of text data. In order to gather data, we will use the Instagram Basic Display API. This API provides the functionality of reading basic data such as comments, hashtags and media files.

There are two fundamental data mining tasks that can be considered in conjunction with Instagram data, namely, text analysis and symbol analysis (emojis). Emojis are a pictorial representation of the facial expression. In addition, their role is mainly pragmatic: emojis give a positive, neutral or negative sense to written sentences by a visual expression (punctuation and letters or pictures). They have been distinguished in three main categories, these sentiment labels can take one of three ordered values (sentiment polarity): negative $\prec$ neutral $\prec$ positive. In sentiment, analysis emojis affect the weight of the given text. Also depending on it, it will change the polarity values of text [2], [3].

For implementation of text processing we used NLTK - Natural Language Toolkit - the most full-featured open-source library of Python, which provides base methods that allow us to classify data into segments (positive or negative) as well as determine the polarity of the input data. In the case of text processing, classification is the task of choosing the correct label for a given text. Since the language in its origin cannot be properly processed by the machine, firstly, the text needs to be split into smaller pieces. For this purpose, NLTK covers libraries for tokenization, parsing, classification, stemming, lemmatization, tagging, and semantic reasoning. Furthermore, with a lot of training data which contains the correct label for each input, the algorithms attempt to incorporate principles of grammar with various natural language processing techniques and mathematical disciplines such as statistics and probability, to train the machine to truly 'understand' the human language.

Since our input data may contain noise - stop words, such as am, in, then we need to first remove all irrelevant words that do not add any meaning to the text. The essence of using NLTK in sentiment analysis is to retrieve 2 parameters for further research: polarity (positive/negative/ neutral) and subjectivity (polar/neutral).

Let us analyze these principles with more detail. The polarity of the word answers the question of whether the word seems to evoke something positive or something negative. However, in this case, polarity must be considered in phrase-level. Some words in preposition may change the contextual polarity of the text to the opposite (e.g not pretty). The expected output of polarity is a number ranging from 0 (absolutely negative) to 1(absolutely positive), which measures the polarity of the phrase. Therefore, with these evaluations, the score polarity of the sentence This cat is not pretty. will be about 0.3 (mostly negative) [4].

Another important task for NLTK is to identify the subjectivity of the text. The concept of this process is to categorize textual information into facts and opinions. Facts are objective expressions about entities, events and their properties. Opinions are usually subjective expressions that are based on people's sentiments, appraisals or feelings. Consequently, the subjectivity will also return a real number from 0 to 1, indicating the score of how subjective/objective the sentence is [5].

After two real numbers - polarity and subjectivity of the text are computed, in order to obtain the final value of polarity, we must transform our crisp data into fuzzy data on the base of the knowledge. This process is called fuzzification. Fuzzy membership function will take these two numerical parameters as input, make its calculations and then report the final number (also from 0 to 1) - score of the prevalence of positive or negative vibes. Fuzzy sets are very helpful there due to its ability to deal with approximate reasoning by allowing gradual membership value assignment.

The provided diagram represents the algorithm of the sentiment analysis report, showing the steps of data processing. When the user sends a post, we first fetch the comments from it, interpreting them as the input data. Then, in order to analyze comments more deeply, we need to divide them into two groups - text and emojis. At that point, using the NLTK tools, we process the

text and get data at the output - subjectivity and polarity indicator (from 0 to 1). And they will be used as input parameters for fuzzy calculations. We will finally get a result of sentiment analysis after all the processes.

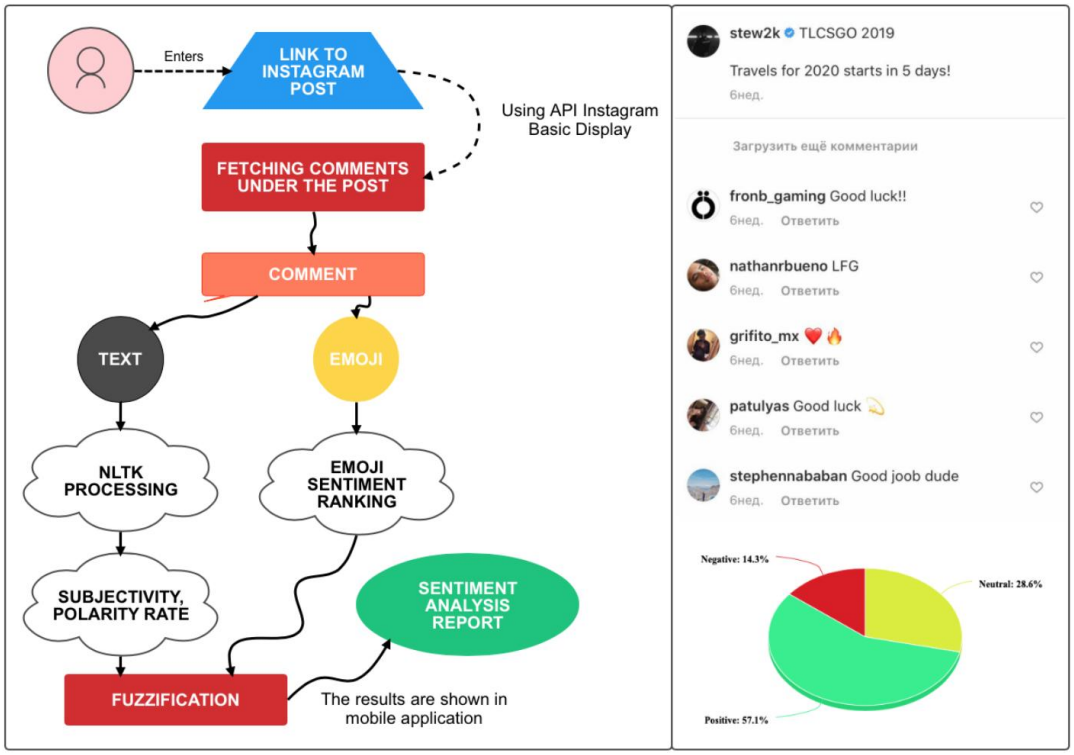

Figure 1 – Proposed model of sentiment analysis

We developed a prototype application that is considered to analyze the comments under the Instagram post. Application helps users in understanding the attitude of people to specific event or opinion expressed in Instagram post. The application has such flow: the user should enter the link of the post he wants to gather statistics. As soon as the user does this, the application downloads all comments related to that post and starts to analyze them. Finally, when data is processed, the application provides the classification to positive and negative ones and show some statistics.

REFERENCES

1. Ronen Feldman. 2013. Techniques and applications for sentiment analysis. Commun. ACM 56, 4 (April 2013), 82–89.
2. Boia M, Faltings B, Musat CC, Pu P. A:) is worth a thousand words: How people attach sentiment to emoticons and words in tweets. In: Intl. Conf. on Social Computing (SocialCom). IEEE; 2013. p. 345–350.
3. Hogenboom A, Bal D, Frasincar F, Bal M, de Jong F, Kaymak U. Exploiting emoticons in sentiment analysis. In: Proc. 28th Annual ACM Symposium on Applied Computing. ACM; 2013. p. 703–710.
4. Wilson T., Wiebe J., Hoffmann P. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. October 2005 p. 347–354.
5. Liu, B. (2010). Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing.

Зиманова Д.А., Садир А.К., Касымов Б.М.
Ғылыми жетекші: Шамои П.С.

Айқын емес жиындар мен логиканы әлеуметтік желілердегі мәтіндердің реңкілігін талдау үшін қолдану

Аңдатпа. Бұл мақала әлеуметтік желілердегі пікірлерге талдау жасау үшін айқын емес логиканы қолдануға арналған. Жұмыстың ерекшелігі - талдау алгоритмін қолдану ғана емес, сонымен қатар iOS платформасында прототиптік қосымшаны жасау болып табылады. Біздің қосымшамыздың мақсатты аудиториясы арасында блогерлер, бизнес иелері, SMM менеджерлері, суретшілер, тіпті мемлекеттік қызметкерлер де бар.

Қазіргі уақытта ғаламторда күн сайын үлкен көлемде ақпарат пайда болады. Алайда, барлық деректер желілік пайдаланушылар үшін ашық болғанымен, оларды өңдеу мүмкін емес. Біздің қосымшаны қолдана отырып, қолданушы әлеуметтік желілердегі өзекті тақырыптар туралы адамдардың жалпылама пікірін ала алады.

Кілт сөздер: Мәтін реңкілігі, айқын емес логика, айқын емес жиындар, мәтінді талдау, эмодзи талдау.

Зиманова Д.А., Садир А.К., Касымов Б.М.
Научный руководитель: Шамои П.С.

Использование нечетких множеств и логики для сентимент-анализа текста в социальных сетях

Аннотация. Данная статья посвящена использованию нечеткой логики в сентимент-анализе комментариев в социальных сетях. Особенностью работы является то, что задача заключается не только в использовании нового алгоритма для сентимент-анализа, но и в разработке прототипного приложения на платформе iOS. Целевой аудиторией данного приложения являются блогеры, владельцы бизнесов, SMM-менеджеры, артисты и даже правительственные служащие.

В наше время огромное количество информации генерируется в интернете каждый день. Однако все эти неструктурированные данные, хоть и прозрачны для пользователей сети, их все же практически невозможно обработать. Используя наше приложение, пользователь может получить обобщенное мнение людей касательно тем в социальных сетях.

Ключевые слова: сентимент-анализ, нечеткие множества, нечеткая логика, анализ текста, анализ эмодзи.

About authors:

Pakizar S. Shamoi, PhD, professor, Kazakh-British technical university.

Dinara A. Zimanova, 4th year student of Kazakh-British technical university, specialty “Information Systems”.

Aigerim K. Sadir, 4th year student of Kazakh-British technical university, specialty “Computer System and Software”.

Bekdaulet M. Kassymov, 4th year student of Kazakh-British technical university, specialty “Information Systems”.