THE PUBLISHING HOUSE
OF THE ROMANIAN ACADEMY

MEDICINE
Research article

EXPERIMENTED KINETIC ENERGY AS FEATURES FOR NATURAL LANGUAGE CLASSIFICATION

Alexandru DAIA[1],Stelian STANCU[1] and Constantin Ionescu-TÎRGOVISTE[2]

Faculty of Cybernetics Statistics and Economic Informatics, Bucharest, Romania[1]
National Institute of Diabetes, Nutrition and Metabolic Diseases, NC Paulescu, Bucharest, Romania[2]
Corresponding author: Alexandru DAIA: alexandru130586@yandex.com

Accepted August 24, 2018

This article describes various uses of kinetic Energy in Natural Language Processing (NLP) and why Natural Language Processing could be used in trading, with the potential to be use also in other applications, including psychology and medicine. Kinetic energy discovered by great Romanian mathematician Octave Onicescu (1892-1983), allows to do feature engineering in various domains including NLP which we did in this experiment. More than that we have run a machine learning model called xgboost to see feature importance and the features extracted by xgboost where captured the most important, in order to classify for simplicity of reader some authors by their content and type of writing

Keywords: feature extraction, machine learning, Octav Onicescu kinetic energy on random vectors, cryptocurrencies, bitcoin, natural language processing, sentiment analysis.

INTRODUCTION

Kinetic energy for random vectors can be referred to as the analogous of kinetic energy in probability from physics. It is can also be defined as entropy that is used for determining uncertainty while measuring information. However, the correct way of explaining kinetic energy is by picturing it as ½ * m * v^2 of a random vector. Octav Onicescu (1892–1983) discovered Kinetic energy and in his description, this entropy is simple sum of square probabilities.

Case in point, if X is a random vector whose probabilities are p1………pN, its kinetic energy becomes Pi2. The calculation for kinetic energy as presented by Octav Onicescu is not different from the one done in physics only that in this case, the (1/2) * mass, has been eliminated based on the assumption that the probabilities are considered to be floating things and the sum of their mass is less plausible (1 or 0). For instance, if we have a random variable X=1, 1, 1, 1, 3, 5, 3; the kinetic energy can be computed as follows: Random vectors are 1, 3, 5.

Their categories are:
Prob (1) = 3/6 =1/2
Prob (3) = 2/6 =1/3
Prob (5) =1/6

Kinetic energy(X) = "sum of squared probabilities"
= Prob $(1)^2$+ Prob $(3)^2$ + Prob $(5)^2$
= $(1/2)^2 + (1/3)^2 + (1/6)^2$
=0.3888

If X = 1, 1, 1, 1, 1…1, 1. Its kinetic energy (X) = Sum (Prob (1) 2) =1.

This thus means that without diversity, kinetic energy remains perfect in a maximum value of 1. In this case, by making an analogy with atomic nuclei, the example has similar maximum kinetic energy just like the one in which atomic nuclei come close together a situation that results in the release of large amounts of energy as in the phenomenon called nuclear fusion.

If X =1, 2, 3, 4, 5…+ inf kinetic energy
Kinetic(X) = 0

This means that in the case of maximum level of diversity and high level of uncertainty, the kinetic energy decreases to a very low value that is zero or near zero. In this case, the random vector categories are interpreted as resulting from atomic nuclei as is the case in the previous example and thus resulting in high energy which then results in large number of atoms but low energy in the end. The phenomenon is called nuclear fission. Based on the two examples, the kinetic energy is bounded between 1, and 0.

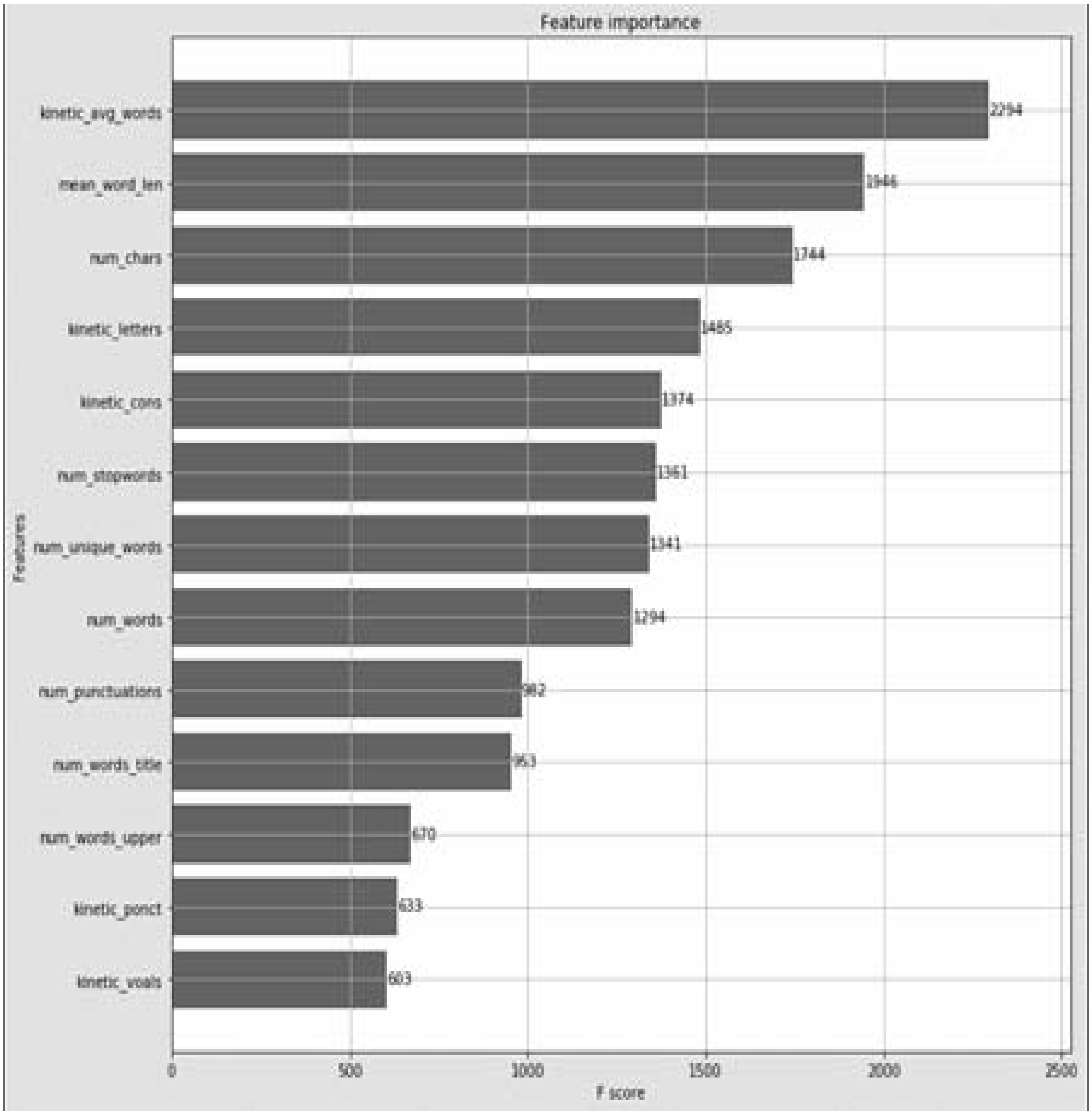

Figure 1. The data from this graphic indicates feature importance of kinetics.

PART 2: THE EXPERIMENT

In this year's Halloween playground competition, you're challenged to predict the author of excerpts from horror stories by Edgar Allan Poe, Mary Shelley, and HP Lovecraft.

The competition is encouraging us to share our insights in the competition's discussion forum and code in Kernels. There are designated prizes that awarded to the authors in these discussions that are particularly valuable to the community which motivate us further. We have conducted an experiment for the Kaggle competition for various features of kinetic energy. We extracted the following features based on the kinetic energy in the Kaggle experiment (Fig. 1).

In our experiments we used the following features.

- Kinetic vowels
- Kinetic consonants
- Kinetic punctuation
- Average of kinetic energy of words
- Kinetic of all letters in the text

KINETIC VOWELS

There is a way of teaching vowels using kinetics. Kinetic vowels are grouped under the articulatory phonetics which is a subsequent category of phonetics in general. These vowels are concern by the transformation of aerodynamic energy to acoustic energy (study of waves in gases, liquids and solids) the aerodynamics in this case refers to the air flow through the vocal tract. Its potential form is air pressure while its kinetic form is the dynamic airflow.

KINETIC CONSONANTS

In kinetic consonants, there is a combination of aerodynamics transformation to acoustic energy. During pronunciation of these consonants, there is energy of motion that is used (kinetic energy) to adjust different parts of the mouth to produce the sounds correctly. The sound energy produced, is also as a result of kinetics. Generally, kinetic motions are responsible for the various adjustments of the mouth shapes to produce these sounds. The modification of the mouth to force a dynamic airflow can be attributed to kinetic energy.

KINETIC PUNCTUATION

In language, punctuations are important as they bring about different meanings to a specific sentence/text or paragraph. Understanding how kinetic energy is utilized can be crucial in punctuating the sentence/text or paragraph. The regulation of the amount of energy produced to propel the air in the tract is as a result of kinetic motions.

AVERAGE OF KINETIC ENERGY OF WORDS

Words have some energy in form of gas molecules. Therefore words can be considered in the bracket of gas molecules. However, with the different forms of energy that words are made, it is possible to use kinetic formulae to calculate the average energy of the words.

KINETIC OF ALL LETTERS IN THE TEXT

In the pronunciation of different letters in a text, it is efficient that the text be broken down into single letters which comprise of vowels and consonants. Each letter is pronounces with different dynamics which involve the movement of the mouth to produce sound energy that produces the pronunciation. All letters in a text poses different potentials in seemingly different ways. It is therefore possible to come up with the kinetic energy of all the letters in a text. To conclude, understanding different languages and trend in the languages, kinetics is essential. These can be as a new method of extracting relevant features in natural language, with a potential application also in medical fields.

CONCLUSION

We have obtained some good features demonstrated as influential for classification task according to the features importance picture in the Figure 1. The classification could be not restricted only to predict author of text but could extend to financial markets or cryptocurrencies in the sense some could use news and this extracted features to try to classify up or down of price considering the impact of some representative news some could have.

These days, according to[7], only 5% of investors make financial gains from crypto news study, this representing another reason why robust features as the ones developed by this experiment will prove that will work well in another demonstration that we are working on currently. According to[8] states that, it is a good thing not to trade your own emotions, but instead get emotions/sentiments from text data.

Also according to[9] Findings, the authors say that “This paper has identified that interaction between media sentiment and the Bitcoin price exists and that there is a tendency for investors to overreact on the news in a short period.” Given only these examples but there are tones out there, we could conclude that researched and done practically. Onicescu Energy to exact features from the text will prove beneficial on a long time horizon.

Judging from the medical perspective, which could also have an impact in the pharmaceutical and financial sector. We are having in a plan to do research using text data from diabetes and other metabolic disorders.

According to[10] the budget allocated for diabetes in the year 2014 was much more significant compared to the year 1994. More exactly poems were written by diabetes and non-diabetes persons to see if we could detect patterns that could discriminate between the two cases, and this makes sense other did it with different approaches and for other diseases.

Recent studies have found distinct differences between the handwriting of patients with Parkinson’s disease and that of healthy people” is concluded in[11]. Or according to[12], this method could be useful in “Handwriting Analysis Diagnoses Major Nervous System Disorders, But Reveals Little Else.”

ACKNOWLEDGEMENT

The autors has nothing to declare related to this article.

REFERENCES

1. Alipour, M., &Mohajeri, A. Onicescu information energy in terms of Shannon entropy and Fisher information densities. An International Journal at the Interface Between Chemistry and Physics, 2012. 110(7), 403-5.
2. Chatzivvas, K., Moustakidis, C., &Panos, C. Information entropy, information distances, and complexity in atoms. The Journal of Chemical Physics. 2005.
3. Simple Feature Engg Notebook - Spooky Author. (2017). Retrieved November 28, 2017, from https://www.kaggle.com/sudalairajkumar/simple-feature-enggnotebook-spooky-author
4. Daia, A. Kinetic things- by DaiaAlexandru. 2016, from https://alexandrudaia.quora.com/Kineticthings-by-Daia-Alexandru
5. Yahya, W., Oyewumi, K., &Sen, K. (2014). Position and momentum informationtheoretic measures of the pseudoharmonic potential. 2014, from https://arxiv.org/abs/1409.7567
6. Rizescu, D., &Avram, V. Using Onicescu's Informational Energy to Approximate Social Entropy. Procedia - Social and Behavioral Sciences, 2014: 377-81.
7. https://www.express.co.uk/finance/city/994622/Bitcoin-price-ripple-cryptocurrency-ethereum-BTC-to-USD-XRP-news.
8. https://news.bitcoin.com/sentiment-analysis-is-the-best-trading-tool-youre-not-using/
9. https://www.independent.co.uk/life-style/gadgets-and-tech/news/cryptocurrency-prices-bitcoin-investment-mood-study-value-a8363866.html.
10. http://www.viatacudiabet.ro/controlul-diabetului/stiri-despre-diabet/numarul-persoanelor-cu-diabet-zaharat-este-in-crestere-53.
11. https://www.jpost.com/Health-and-Science/Handwriting-assessment-can-be-used-for-early-detection-of-Parkinsons-disease-325798.
12. https://www.medicaldaily.com/handwriting-analysis-diagnoses-major-nervous-system-disorders-reveals-little-else-291972.