International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

Design Analysis of Banking Transaction Descriptions via Deep Learning Mechanism

Pallavi D. Bankar, Prof. Anil V. Deorankar

Computer Science and Engineering, Government College of Engineering, Amravati, Maharashtra, India

Article Info
Volume 8, Issue 5
Page Number : 272-277

Publication Issue
September-October-2021

Article History
Accepted : 01 Oct 2021
Published : 07 Oct 2021

ABSTRACT

A non-invasive technique using knee joint vibroarthographic (VAG) signals can be used for the early diagnosis of knee joint disorders. Among the algorithms devised for the detection of knee joint disorders using VAG signals, algorithms based on entropy measures can provide better performance. In this work, the VAG signal is preprocessed using wavelet decomposition into sub band signals. Features of the decomposed sub bands such as approximate entropy, sample entropy & wavelet energy are extracted as a quantified measure of complexity of the signal. A feature selection based on Principal Component Analysis (PCA) is performed in order to select the significant features. The extracted features are then used for classification of VAG signal into normal and abnormal VAG using support vector machine. It is observed that the classifier provides a better accuracy with feature selection using principal component analysis. And the results show that the classifier was able to classify the signal with an accuracy of 82.6%, error rate of 0.174, sensitivity of 1.0 and specificity of 0.888.

Keywords : Vibroarthrography, Wavelet decomposition, Feature extraction, Principal Component Analysis, support Vector Machine.

I. INTRODUCTION

The number of financial institutions introducing and growing their electronic banking products and services is increasing all the time. Each financial institution's long-term goal is to maintain existing clients while also acquiring new customers who are not now customers. Know Your Customer (KYC) is a regulatory-approved structure that is used in manual banking to restrict the monetary behaviour of customers as well as their perception of the organisation. There are locations where investor clients are exposed to a high level of risk, while other areas are exposed to medium risk, and the remaining areas have a safe haven where they can invest. In the current state of affairs, credit risk for a counterparty can be classified into two categories: quantitative and subjective factors. Despite the fact that there are numerous existing frameworks on customer maintenance, as well as client steady loss situations in banks, these comprehensive strategies continue to provide a clear and defined strategy to deal with dispensing credit in the commercial area [1] [2, 3].

Because of the large number of publicly available advanced text sources, automated text classification has become a popular research area. In a wide range of applications, text classification is beneficial, such as web searching, opinion mining [2], and event detection. The use of target-dependent sentiment analysis on Twitter has drew increased attention to the field of exploration. The majority of previous work is based on linguistic structure, such as artificial parse trees, which are susceptible to noise when dealing with informal material such as tweets.

When dealing with the instabile development of client-produced interactive media assets in the vast information period, bridging the semantic gap between low-level highlights and significant level semantics in various methods of information is both a fundamental and needless concern. In this study [3], the authors take advantage of a well-known type of information known as collaborative tags to overcome this issue. In particular, they distinguish between essential level ideas and then build ontologies from the cooperative labels, as seen in the following figure. The ontologies that have been built can be used to organise and index media information that has been created by clients. Existing exploration does not have a guiding principle that directs the extraction of philosophy from a human point of view.

In recent years, a few frameworks have been developed that allow clients to evaluate and clarify assets. These frameworks believe that combining the client's rating and tags provides an approach to distinguishing between favourable and bothersome tags. Using momentum work on client profiling for tailored search in community-oriented tagging frameworks, we identify and elaborate on the limitations of the work presented in this study. Another staggered client profile strategy, incorporating tags and ratings to achieve tailored search, is then proposed [4]. This model can show both the client's approval as well as the client's dissatisfaction.

Even after decades of research, distinguishing between similar languages remains a difficult problem–no known system has achieved flawless performance. The models described in this research once again demonstrate that classical models, such as support vector machines (SVMs), outperform deep learning techniques on this problem [5, 6].

Native Language Identification (NLI) is the job of intuitively recognising an individual's native language (L1) based on their language production in a learnt language, which is dependent on their language creation in a native language. It is often described as a classification problem in which the layout of L1s has been deduced from previous knowledge. Two previous collaborative tasks on NLI were organised, with the goal of distinguishing the L1 of students of English based on writings (2013) and oral reactions (2016) they provided during a standardised appraisal of scholastic English proficiency. The inputs from the two previous problems are combined in an intriguing way in the 2017 common task [6].

II. LITERATURE SURVEY

A. Customer Analysis

Given the significance of customers as the most important resources of associations, customer maintenance is by all accounts a fundamental, essential prerequisite for any association. Banks are no special case for this standard. The serious climate within which electronic banking administrations are given by various banks increases the need of customer maintenance. Being founded on existing information advancements which permit one to gather information from associations' data sets, information mining introduces a useful asset for the extraction of knowledge from gigantic measures of information. In this examination [7], the decision tree procedure was applied to construct a model incorporating this knowledge. Bank directors can recognize churners in

future using the consequences of decision tree. They ought to give a few procedures to customers whose highlights are getting bound to churner's highlights.
This investigation [8] reveals the impact of the length, recency, frequency, monetary, and profit (LRFMP) customer esteem model in a coordination's organization to anticipate customer beat. This special context has valuable business suggestions contrasted with the main stream customer stir examines where individual customers (instead of business customers) are the main core interest.

Analyzing customer input is the most ideal approach to channelize the information into new marketing techniques that advantage business visionaries just as customers. Hence a computerized framework which can examine the customer conduct is in incredible interest. Clients may compose inputs in any language, and subsequently mining suitable information regularly gets intractable. Particularly in a conventional element based regulated model, it is hard to fabricate a nonexclusive framework as one needs to comprehend the concerned language for finding the pertinent highlights [9].

B. Personal Finance Management

The situated decision support system (SDSS) model is a type of DSS that maintains close links with the target environment and has capabilities for sensing, monitoring, decision support, and limited decision making, action generation, and implementation. Yet, however a conventional portrayal of SDSS had been given, no experimental test has been made to demonstrate its worth. Creator [10] performed tests using human subjects to test a SDSS model. Personal finance management was chosen as an application domain, and the design and implementation of the SDSS prototype is discussed.

C. Open Banking European Regulation

The paper breaks down [11] these four columns and recommends that together they will underpin the eventual fate of advanced financial administrations in Europe, and – together – will drive a Big Bang progress to information driven finance. These seemingly random columns together reinforce an emerging environment which expects to advance an equilibrium among a scope of some of the time conflicting targets, including foundational hazard, information security and security, productivity, and customer assurance. The European path to digitization is based on four pillars:

1. Extensive reporting requirements to control systemic risk and change financial sector behaviour;
2. Strict data protection rules;
3. Open banking to enhance competition; and
4. A legislative framework for digital identification.

In this line, the Second Payments Services Directive6 (PSD 2) empowers customers to make their banking data available to third parties such as *FinTech* companies. In essence it paves the way for new banking products and services, by promoting competition without compromising security.

D. Text Classification

Ongoing analysts have utilized convolutional neural networks or intermittent neural networks for text classification inspired by the observable achievement of deep learning. Notwithstanding, the greater part of their models depend on single network. This paper [12] combines the convolutional neural network (CNN) that is worthwhile in extracting neighborhood highlights with the intermittent neural network, or all the more explicitly, the Long Short-Term network, that has amazing memory, and proposes a Convolutional Recurrent Neural Network model for text classification.

Sentiment analysis of short texts, for example, single sentences and Twitter messages is challenging a direct result of the restricted contextual information that they typically contain. Successfully solving this errand requires techniques that combine the small text content with earlier knowledge and utilize something beyond pack of-words. In this work [13] propose another deep convolutional neural network that misuses from character-to condemn level information to perform sentiment analysis of short texts.

Transductive classification is a valuable method to classify texts when labeled training models are insufficient. A few algorithms to perform transductive classification considering text assortments addressed in a vector space model have been proposed [14]. Nonetheless, the utilization of these algorithms is impractical in functional applications because of the independence presumption among instances or terms and the drawbacks of these algorithms. Network-based algorithms come up to stay away from the drawbacks of the algorithms dependent on vector space model and to improve transductive classification.

By and large, word use is distinctive depending on the context of reports, and it is sensible for words to coincide with different words sharing similar context in the two archives having a comparative meaning. At the end of the day, the co-event esteem between words in the various archives indicates a closeness of these records which have similar context or substance. By using a combination of words increases tokens, and [15] expect that we will actually want to get a handle on sensitive highlights like context.

The [16] present ALL-IN-1, a simple model for multilingual text classification that does not require any parallel data. It is based on a traditional Support Vector Machine classifier exploiting multilingual word embedding's and character n-grams.

In this Article [17] describe a novel system that combines Natural Language Processing techniques with Machine Learning algorithms to classify banking transaction descriptions for personal finance management, a problem that was not previously considered in the literature. They trained and tested that system on a labeled dataset with real customer transactions that will be available to other researchers on request. Motivated by existing solutions in spam detection, also propose a short text similarity detector to reduce training set size based on the Jaccard distance.

III. PROPOSED METHODOLOGY

A. Dataset Description

The dataset downloaded from https://data.world/lpetrocelli/retail-banking-demo-data/

The entries of the dataset have the following attributes:

1) ID: a unique numeric identifier.

2) Description: the BT short-text description.

3) Amount: the amount in euros of the BT, either positive (income) or negative (expense).

4) Date: the date when the BT occurred.

B. Proposed System

Following fig. 1 shows the proposed system architecture. System uses Banking transaction having short text description as input dataset. Our proposed approach has four stages, as described in Figure 1: (1) Preprocessing (data cleaning), (2) Similarity Finding (Cosine Similarity) (3) Modeling (Features extraction and model training), and (4) classification (BERT). Proposed approach uses BERT Machine Learning algorithm for classification, by distinguishing the end-to-end machine learning model from multi-label knowledge.

Preprocessing

Text descriptions may still contain useless information that may affect text classification performance. To improve the classification performance, the input dataset is clean using stemming and stopwords removal processes.

Similarity Finding

Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. It is often used to measure document similarity in text analysis. Cosine similarity is computed using the following formula:

$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Values range between -1 and 1, where -1 is perfectly dissimilar and 1 is perfectly similar.

Classification

BERT and other Transformer encoder architectures have been wildly successful on a variety of tasks in NLP (natural language processing). They compute vector-space representations of natural language that are suitable for use in deep learning models. BERT models are usually pre-trained on a large corpus of text, then fine-tuned for specific tasks.

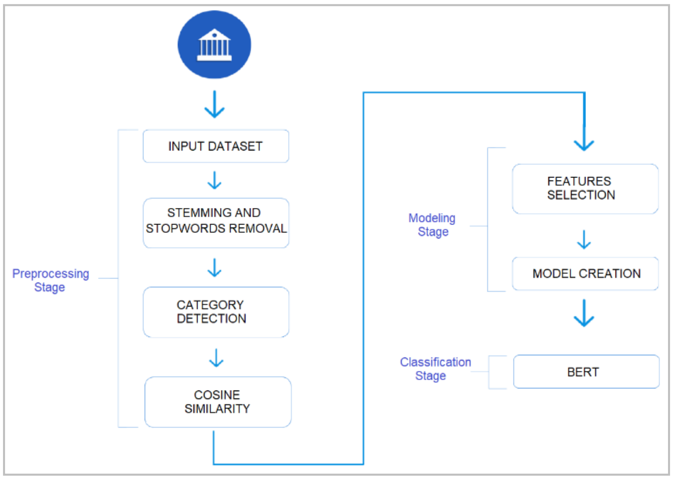

Fig. 1 Proposed System Architecture

IV. CONCLUSION

In this paper, we present a unique framework for categorising financial transactions based on short textual descriptions. Comparing our framework to existing algorithms for brief text classification, we find that it has two distinct advantages: For the purpose of avoiding mismatches, it compares the similarities of short texts from the perspective of their concepts. ii) Less training data is required to learn the concept model for each class because only a small number of phrases are capable of representing a single concept. Because of these characteristics, it is well suited for lightweight systems that must deal with brief texts and have strict requirements for fast learning and word altering adaptations. The categorization work was carried out with the help of the BERT classifier.

V. REFERENCES

[1]. M. R. Islam and M. A. Habib, ``A data mining approach to predict prospective business sectors for lending in retail banking using decision tree,'' CoRR, vol. abs/1504.02018, no. 2, pp. 13-22, 2015.

[2]. D.-T. Vo and Y. Zhang, ``Target-dependent Twitter sentiment classification with rich automatic features,'' in Proc. IJCAI, 2015, pp. 1347-1353.

[3]. Y. Cai, W.-H. Chen, H.-F. Leung, Q. Li, H. Xie, R. Y. K. Lau, H. Min, and F. L. Wang, ``Context-aware ontologies generation with basic level concepts from collaborative tags,'' Neurocomputing, vol. 208, pp. 25-38, Oct. 2016.

[4]. Q. Du, H. Xie, Y. Cai, H.-F. Leung, Q. Li, H. Min, and F. L. Wang, ``Folksonomy-based personalized search by hybrid user profiles in multiple levels,'' Neurocomputing, vol. 204, pp. 142-152, Sep. 2016.

[5]. M. Medvedeva, M. Kroon, and B. Plank, ``When sparse traditional models outperform dense neural networks: The curious case of

discriminating between similar languages,'' in Proc. VarDial, 2017, pp. 156-163.

[6]. S. Malmasi, K. Evanini, A. Cahill, J. Tetreault, R. Pugh, C. Hamill, D. Napolitano, and Y. Qian, ``A report on the 2017 native language identification shared task,'' in Proc. 12th Workshop Innov. Use NLP Building Educ. Appl., 2017, pp. 62-75.

[7]. A. Keramati, H. Ghaneei, and S. M. Mirmohammadi, ``Developing a prediction model for customer churn from electronic banking services using data mining,'' Financial Innov., vol. 2, no. 1, p. 10, Dec. 2016.

[8]. K. Chen, Y.-H. Hu, and Y.-C. Hsieh, ``Predicting customer churn from valuable B2B customers in the logistics industry: A case study ,'' Inf. Syst. e-Bus. Manage., vol. 13, no. 3, pp. 475-494, Aug. 2015.

[9]. B. Plank, ``All-in-1: Short text classification with one model for all languages,'' in Proc. Int. Joint Conf. Natural Lang. Process. (Shared Task 4), Dec. 2017.

[10]. R. Vahidov and X. He, ``Situated DSS for personal finance management: Design and evaluation,'' Inf. Manage., vol. 46, no. 8, pp. 453-462, Dec. 2009.

[11]. D. A. Zetzsche, D. W. Arner, R. P. Buckley, and R. H. Weber, ``The future of data-driven finance and RegTech: Lessons from EU big bang II,'' Australas. Legal Inf. Inst., Univ. New SouthWales Law Res. Ser., Sydney, NSW, Australia, Tech. Rep., 2019.

[12]. R. Wang, Z. Li, J. Cao, T. Chen, and L. Wang, ``Convolutional recurrent neural networks for text classification,'' in Proc. AAAI, vol. 333, 2015, pp. 2267-2273.

[13]. C. dos Santos and M. Gatti, ``Deep convolutional neural networks for sentiment analysis of short texts,'' in Proc. 25th Int. Conf. Comput. Linguistics (COLING), 2014, pp. 69-78.

[14]. R. G. Rossi, A. D. A. Lopes, and S. O. Rezende, ``Optimization and label propagation in bipartite heterogeneous networks to improve transductive classification of texts,'' Inf. Process. Manage., vol. 52, no. 2, pp. 217-257, Mar. 2016.

[15]. S. Temma, M. Sugii, and H. Matsuno, ``The document similarity index based on the Jaccard distance for mail filtering,'' in Proc. 34th Int. Tech. Conf. Circuits/Syst., Comput. Commun. (ITC-CSCC), Jun. 2019, pp. 1-4.

[16]. D. Gupta, P. Lenka, H. Bedi, A. Ekbal, and P. Bhattacharyya, ``IITP at IJCNLP-2017 task 4: Auto analysis of customer feedback using CNN and GRU network,'' in Proc. IJCNLP, 2017, pp. 184-193.

[17]. Silvia García-Méndez, Milagros Fernández-Gavilanes, Jonathan Juncal-Martínez, Francisco Javier González-Castaño, And Óscar Barba Seara, "Identifying Banking Transaction Descriptions via Support Vector Machine Short-Text Classification Based on a Specialized Labelled Corpus", IEEE, VOLUME 8, April 13, 2020.

Cite this article as :

Pallavi D. Bankar, Prof. Anil V. Deorankar, "Design Analysis of Banking Transaction Descriptions via Deep Learning Mechanism", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 5, pp. 272-277, September-October 2021.

Journal URL : https://ijsrst.com/IJSRST218549