Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

Experiment No. 1
Review of Deep Learning techniques
Date of Performance: 25/07/23
Date of Submission: 01/08/23

Experiment – 01

Paper 1 – Linguistic Anomaly Detector

Introduction:

Identifying anomalous sentiment patterns, or unique textual characteristics of such patterns in a set of textual data, is known as anomaly detection in text mining. The identified anomalies could be the result of abrupt changes in decision-making for text classification problem types. If these abnormalities go unnoticed or are improperly handled, consequences could result, such as a poor performance in text classification systems Many web applications can benefit from the capacity to spot anomalies in streams of text data. It can be used, for instance, to identify significant events from Twitter streams, fraudulent email exchanges, and even inaccurate descriptions in maintenance logs, authors tried to provide a solid and comprehensive overview of on anomaly detection.

Problem Statement:

The paper addresses the need to distinguish between normal and abnormal textual patterns and highlights the complexities of distinguishing novelties from outliers. By providing examples of applications in various domains, the problem statement illustrates the real-world significance of the anomaly detection task. The author's emphasis on the potential consequences of missed anomalies enhances the reader's understanding of the problem's implications.

Model:

The author introduces key components of the model, including Autoencoders, Sentence BERT (SBERT), and logistic regression. The use of visual aids, such as diagrams, helps the reader grasp the architecture and flow of the model. By explaining the architecture of the Autoencoder, including encoder and decoder components, and integrating SBERT for creating rich sentence embeddings, the section effectively conveys how the model transforms textual data into numerical representations. Additionally, the author's explanation of how the AE is trained, reconstruction errors are calculated, and the combination with logistic regression takes place provides a clear understanding of the entire process.

Technologies:

The technologies section provides a valuable overview of the tools and methods used in the paper. It highlights the importance of deep learning architectures, transformers, and sentence embeddings. By mentioning specific models like BERT and SBERT, the author ensures that the reader understands the foundations of the proposed approach. The mention of specific techniques, such as convolutional layers, mean pooling, and logistic regression.

Conclusion:

The author effectively summarizes the advantages of the proposed approach and its potential impact on reducing false positive rates. The mention of possible future work involving the application of the model to other datasets adds a forward-looking perspective.

Paper 2 - TextCharm: Anomaly Detection at the Character Level

Introduction:

It effectively highlights the significance of anomaly detection in various application domains and explains the need for accurate identification of anomalies. It presents the context of the problem by mentioning its importance in fields like information security, e-commerce, and medicine. The introduction also establishes the need for a data-driven approach to anomaly detection, particularly in cases where supervised learning is impractical due to limited labeled data. It introduces the concept of character-level representation for text data and sets the stage for the research problem.

Problem Statement:

The problem statement articulates the challenges of anomaly detection and interprets the nature of anomalies in text data. It defines different types of anomalies such as unintended typos, fraud, and intended modifications, providing clear examples for each. The problem statement sets the foundation for the research's focus on developing a fully data-driven model to detect anomalies in textual data sets, addressing real-world business cases.

Model:

The model section elaborates on the proposed approach for character-level anomaly detection. It explains the character representation technique, highlighting how text instances are converted into numerical features based on encoded characters. The section goes on to detail the design of the ensemble model, which combines various anomaly detection classifiers such as PCA, COPOD, HBOS, LODA, CBLOF, and IForest. Hyperparameters for each algorithm are provided. The model's architecture, standardization of anomaly scores, and use of ensemble methods (AOM and MOA) are all described in a structured manner.

Technologies:

The technologies section outlines the tools and techniques employed in the research. It mentions the use of Python for model development and incorporation into the MobyDQ framework. It also introduces the sklearn library for certain algorithms and the pyod library for others. The section briefly touches on data preprocessing, such as character quantization and scaling using the MaxAbs scaler.

Conclusion:

The conclusion effectively summarizes the research's key findings and contributions. It highlights the practical implications of the developed anomaly detection model, particularly in maintaining data consistency and quality for business users. The conclusion also suggests future directions for research, including exploring more base algorithms and adapting the model for mixed-type data sets.

Paper 3 - CommentSense: Emotion-Powered Sarcasm Detection

Introduction:

This paper embarks on a journey to explore the intricate realm of sarcasm detection within online comments by intertwining it with the nuances of emotions. It underscores the significance of emotion and sarcasm analysis in contemporary marketing strategies. The authors lay the groundwork for a system that delves into the dual aspects of grouping posts based on emotions and sentiments while concurrently identifying the presence of sarcasm.

Problem Statement:

The crux of the challenge lies in effectively detecting sarcasm in online comments while taking into account the emotional and sentiment-laden context. The authors accentuate the growing importance of amalgamating emotion and sarcasm analysis to refine marketing strategies by gauging public sentiment towards brands.

Model:

The heart of the proposed solution is a comprehensive system that proficiently groups posts based on emotions, sentiments, and sarcasm. This system amalgamates sentiment and emotion identification modules, creating a holistic approach. It harnesses lexical databases, sentiment scores, and diverse sarcasm detection algorithms. The paper further explicates the intricate sentiment and emotion identification process, which involves evaluating word scores, utilizing lexical databases such as WordNet and SentiWordNet, and deploying sarcasm detection algorithms like Emoticon sarcasm detection, Hybrid sarcasm detection, Hashtag Processing, and Interjection Word Start (IWT).

Technologies:

The paper draws upon a plethora of technologies to realize its ambitious system, including lexical databases such as WordNet and SentiWordNet. It also incorporates an array of sarcasm detection algorithms, including Emoticon sarcasm detection, Hybrid sarcasm detection, Hashtag Processing, and Interjection Word Start (IWT).

Conclusion:

This paper presents an innovative and holistic approach that amalgamates emotion and sarcasm analysis within the realm of online comments. However, it leaves room for improvement in terms of providing a more comprehensive performance evaluation and dataset details. While this approach holds substantial promise for refining marketing strategies and understanding online sentiment, further research and validation are indispensable to solidify its applicability and effectiveness.