**Identifying Hot Topic Trends in Streaming Text Data Using Sequential Evolution Model Based on Distributed**

**ABSTRACT**

Hot topic trends have become increasingly important in the era of social media, as these trends can spread rapidly through online platforms and significantly impact public discourse and behavior. As a result, the scope of distributed representations has expanded in machine learning and natural language processing. As these approaches can be used to effectively identify and analyze hot topic trends in large datasets. However, previous research has shown that analyzing sequential periods in data streams to detect hot topic trends can be challenging, particularly when dealing with large datasets. Moreover, existing methods often fail to accurately capture the semantic relationships between words over different time periods, limiting their effectiveness in trend prediction and relationship analysis. This paper aims to utilize a distributed representations approach to detect hot topic trends in streaming text data. For this purpose, we build a sequential evolution model for a streaming news website to identify hot topic trends in streaming text data. Additionally, we create a visual display model and knowledge graph to further enhance our proposed approach. To achieve this, we begin by collecting streaming news data from the web and dividing it chronologically into several datasets. In addition, word2vec models are built in different periods for each dataset. Finally, we compare the relationship of any target word in sequential word2vec models and analyze its evolutionary process. Experimental results show that the proposed method can detect hot topic trends and provide a graphical representation of any raw data that cannot be easily designed using traditional methods.

**INTRODUCTION**

Detecting hot topic trends in real-time is critical in many fields, including marketing, technology, finance, and politics. However, traditional approaches to trend analysis often fall short when it comes to understanding complex and nuanced language use in a continuous stream of data. This is where distributed representation models, such as word2vec come in. Word2Vec allows grouping similar words together and implementing learning algorithms to improve performance on natural language processing tasks [1]. The model has attracted much attention due to its ability to construct the semantic context of words [2,3]. It contains many algorithms and functions and can be implemented in Java, C, and Python. In short, word2vec is a tool used for computing the vector representation of words. It inputs value as text and gives output as word vectors. Although the usage of distributed representation models for creating embeddings is widespread, many unanswered questions remain about the factors that influence its results and its true capabilities [4,5]. These models can efficiently capture the semantic and syntactic relationships between words and phrases, allowing for more accurate and precise trend analysis. In particular, the use of distributed representation models in a distributed computing environment can enable real-time processing of massive amounts of data, making it possible to detect and respond to emerging trends faster than ever before. Therefore, developing and applying distributed representation models for trend analysis is an area of growing importance and interest.

Some of the current issues in hot topic trend detection include the difficulty in handling large amounts of data, as well as the challenge of detecting subtle shifts in language use and topic evolution over different time spans. Different areas of application such as bioinformatics, data mining, speech recognition, remote sensing, multimedia, text detection, localization, and others, require different techniques to be utilized. Therefore, there is no single technique that can be applied universally across all these areas [6]. Understanding the trends in software engineering [7] emphasized the importance of deeper analysis and a systematic approach. In [8] the objective of the study was to examine the research trends in Science, Technology, Engineering, and Mathematics (STEM) education, while [9] gives importance to describing the fields of study and trends in computational thinking (CT), and [10] detected IOT Bot net in 5G Core Network. In [11] the author aimed to hidden discover topics and trends within historical incident reports of the Air Traffic Control (ATC) system. The research work of [12] introduces Bengal-BERT, a monolingual BERT model designed for the Bengal language. Additionally, many traditional methods for analyzing language and identifying topics rely on handcrafted features and rule-based approaches, which can be time-consuming and may not generalize well to different datasets. Furthermore, many of these methods may not be suitable for real-time processing, which can be important for applications such as social media monitoring and sentiment analysis. Therefore, there is a need for new, more efficient, and scalable methods for analyzing language use and detecting hot topic trends in data streams.

The field of natural language processing has experienced significant advancements in recent years, with the rise of machine learning techniques and the availability of large datasets. In previous times, users were required to communicate their requirements and intentions by composing a well-defined document in everyday language [13]. With the growth of social media platforms and online news websites, analyzing text data has become crucial in understanding public opinion, and consumer behavior, and predicting future trends. Streaming text data is becoming more prevalent, and traditional batch processing techniques are no longer sufficient to handle the high volume and variability of this data. One key challenge in this domain is identifying hot topic trends in real-time, as they emerge and evolve. Traditional methods for trend detection often rely on handcrafted features or require significant human intervention, making them slow and prone to errors. In contrast, machine learning techniques based on distributed representations have shown great promise in automatically learning patterns and relationships in large volumes of textual data. Therefore, there is a need for simultaneous analysis of streaming text data to capture evolving trends and patterns.

In this paper, we aim to address this need by proposing a novel approach for detecting hot topic trends in streaming news data using a sequential evolution model based on distributed representations. The approach has the potential to provide valuable insights to market analysts, news agencies, and researchers in various fields. The ability to analyze trends in large volumes of text data is important for a wide range of applications, including trend detection in social media and news analysis. However, existing methods for trend analysis are often limited by the lack of robust techniques for analyzing the relationships between target words. To address this problem, we propose a novel NSEM for analyzing the difference in sequential periods between different data sets. Our approach uses word2vec models to analyze the semantic relationships between words for each dataset separately and identifies trends by comparing these models over periods. This approach allows us to accurately identify trends, provide a data visualization model, and create a knowledge graph of raw streaming news data.

**SYSTEM ANALYSIS**

**EXISTING SYSTEM:**

Since Mikolov et al. [1-3] introduced the word2vec model, many studies have applied the model to learn word embeddings. However, this work focuses on a different application of the word2vec and implements the model to analyze the trend relationships between target words. Additionally, our work introduces a novel model called NSEM to analyze the difference in sequential periods between different data sets, a visual display model, and a knowledge graph to represent trends. Thus, our work is introducing new contributions and applications in the field. Dynamic topicmodeling was presented by [14], where a family of probabilistic time series models is developed to analyze the temporal evolution of topics within extensive documentcollections. The proposed approach involves utilizing state space models on the natural parameters of multinomial distributions that represent the topics. However, our work differs from dynamic topic modeling in several key aspects. While both approaches aim to capture the temporal dynamics of the topics, they employ different methodologies and have distinct strengths.

Dynamic topic modeling (DTM) is a probabilistic modeling technique that enables the discovery of topics that evolve over time in a given document collection. It considers the temporal ordering of documents and captures the changing prevalenceof topics over time. DTM can be effective in identifying topic shifts, tracking topic evolution, and revealing latent thematic patterns in textual data. On the other hand, our proposed NSEM approach focuses on the detection of hot topic trends specifically. We leverage the distributed representation approach, such as word2vec, to capture semantic relationships and patterns within the data. This enables us to identify and analyze trends that gain prominence or decline during diverse time periods. It is important to note that dynamic topic modeling can be a valuable technique in various applications, particularly when the objective is to analyze the evolution of topics over time at a more granular level. In contrast, our focus is on identifying and understanding the trends that are currently popular or gaining traction within a given dataset. Overall, while both dynamic topic modeling and our proposed approach aim to capture temporal dynamics, they differ in their methodologies, objectives, and specificcontributions to the field of trend detection. Each approach has its strengths and can be applied depending on the specific research goals and requirements.

**Disadvantages:**

* + 1. The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets to detect Identifying Hot Topic Trends.
    2. Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.
    3. Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

**PROPOSEDSYSTEM:**

The paper presents several contributions toward trend detection and analysis. Firstly, we explore the use of distributed representations as a promising approach for trend detection, which allows for comprehensive analysis of data by capturing semantic relationships between topics. Moreover, we propose the News Sequential Evolution Model (NSEM) which uses distributed representations to analyze trend relationships between words in large datasets. Additionally, we introduce a unique chronological analysis of trends and relationships between words in different datasets, which considers how they change during different time periods for a more accurate understanding of their evolution. Furthermore, our visualization display model leverages data visualization theory to identify hot topic trends across various domains, providing valuable insights for decision-making and staying ahead of the competition. Finally, we present a word2vecbased knowledge graph representation technique that contributes to identifying patterns and relationships in streaming news data, enabling applications such as question answering systems and recommendation engines. To the best of our knowledge, this paper is the primary work in identifying topic trends from stream text news based on distributed representations. These contributions could potentially have significant implications for a range of applications in the field.

**Advantages:**

* The paper proposes a methodology and toolbox for analyzing streaming data sources, providing users with insights into trending topics related to their products.
* The proposed method has shown promising results in identifying relevant and significant trends for technical words. The analysis of the trends related to "LG" and "Apple" demonstrates how our method effectively filters out low-value trends and captures the relationship between different words over different time periods.

**SYSTEM REQUIREMENTS**

**HARDWARE REQUIREMENTS:**

* System :i3
* HardDisk : 40 GB.
* FloppyDrive : 1.44 Mb.
* Monitor :15 VGAColour.
* Mouse :Logitech.
* Ram :512 Mb.

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 7 Ultimate.
* Coding Language : Python.
* Front-End : Python.
* Back-End : Django-ORM
* Designing : Html, css, javascript.
* Data Base : MySQL (WAMP Server)

**MODULES**

1. Data Coordinator
2. Distant User

**SYSTEM ARCHITECTURE:**

