



Contribution ID: 125

Type: **Research Presentation (30 minutes)**

An N-gram Approach to Identifying Lexical Features in Fake Financial News

Sunday, 19 May 2024 11:30 (30 minutes)

While financial reports play an important role in guiding investment decisions, the research landscape pertaining to fake news detection in the financial sector remains underexplored with most existing literature primarily focusing on false political and social news identification or on the impact of fake financial news on investor attention and market reaction (Zhang & Liu, 2023). To address this gap and seek to enhance the limitation of pure machine learning algorithms, the present study proposed applying the n-gram approach in natural language processing (NLP). Through investigating word-based 4-grams (i.e., quadgrams) and scrutinizing forms and functions of these lexical strings in the fake financial news articles obtained from the Securities and Exchange Commission (SEC) crackdown of stock promotion schemes in the United States in 2017, the present study identified 18 high-frequency quadgrams. Based on the taxonomy devised by Biber and his colleagues (1999), the findings showed that (1) the 18 quadgrams were dominated by PP-based expressions in terms of structures and (2) functionally, referential quadgrams constituted over 75% of the total tokens and types, followed by an additional subcategory (i.e., subject-specific quadgrams) which contained medical or financial vocabulary. It is hoped that through uncovering n-grams alongside qualitative structural and functional analyses, the findings of this study would not only contribute to the detection of illegitimate news but also draw the attention of consumers or readers to some implicit deceptive linguistic patterns in financial news reports.

Is this a sponsored session?

Keywords

N-gram, natural language processing (NLP), fake news, financial news

Primary author: CHENG, Su-han

Presenter: CHENG, Su-han

Session Classification: DN 403

Track Classification: Artificial Intelligence CALL: AI & Ethics, Access, Equality, Diversity, and Inclusion