World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:17, No:11, 2023

An Unsupervised Domain-Knowledge Discovery Framework for Fake News Detection

Authors : Yulan Wu

Abstract : With the rapid development of social media, the issue of fake news has gained considerable prominence, drawing the attention of both the public and governments. The widespread dissemination of false information poses a tangible threat across multiple domains of society, including politics, economy, and health. However, much research has concentrated on supervised training models within specific domains, their effectiveness diminishes when applied to identify fake news across multiple domains. To solve this problem, some approaches based on domain labels have been proposed. By segmenting news to their specific area in advance, judges in the corresponding field may be more accurate on fake news. However, these approaches disregard the fact that news records can pertain to multiple domains, resulting in a significant loss of valuable information. In addition, the datasets used for training must all be domain-labeled, which creates unnecessary complexity. To solve these problems, an unsupervised domain knowledge discovery framework for fake news detection is proposed. Firstly, to effectively retain the multidomain knowledge of the text, a low-dimensional vector for each news text to capture domain embeddings is generated. Subsequently, a feature extraction module utilizing the unsupervisedly discovered domain embeddings is used to extract the comprehensive features of news. Finally, a classifier is employed to determine the authenticity of the news. To verify the proposed framework, a test is conducted on the existing widely used datasets, and the experimental results demonstrate that this method is able to improve the detection performance for fake news across multiple domains. Moreover, even in datasets that lack domain labels, this method can still effectively transfer domain knowledge, which can educe the time consumed by tagging without sacrificing the detection accuracy.

Keywords : fake news, deep learning, natural language processing, multiple domains
Conference Title : ICSLP 2023 : International Conference on Speech and Language Processing
Conference Location : San Francisco, United States
Conference Dates : November 06-07, 2023