A Credibility Analysis System for Assessing Information on Twitter

ABSTRACT

Information credibility on Twitter has been a topic of interest among researchers in the fields of both computer and social sciences, primarily because of the recent growth of this platform as a tool for information dissemination. Twitter has made

it increasingly possible to offer near-real-time transfer of information in a very cost-effective manner. It is now being used as a source of news among a wide array of users around the globe. The beauty of this platform is that it delivers timely content in a tailored manner that makes it possible for users to obtain news regarding their topics of interest. Consequently, the development of techniques that can verify information obtained from Twitter has become a challenging and necessary task. In this paper, we propose a new credibility analysis system for assessing information credibility on Twitter to prevent the proliferation of fake or malicious information. The proposed system consists of four integrated components: a reputation-based component, a credibility classifier engine, a user experience component, and a feature-ranking algorithm. The components operate together in an algorithmic form to analyze and assess the credibility of Twitter tweets and users. We tested the performance of our system on two different datasets from 489,330 unique Twitter accounts. We applied 10-fold cross-validation over four machine learning algorithms. The results reveal that a significant balance between recall and precision was achieved for the tested dataset.

**EXISTING SYSTEM**

* Pal and Scott [9] took a different approach to studying credibility on Twitter: they sought to show how name value bias affects the judgments of microblog authors. In this study, the author showed the correlation between name value bias and the number of followers. A similar study by Morris et al. [14] discussed how users perceive tweet credibility. They conducted a survey that showed a disparity in the features used by users to assess credibility and those that are shown by search engines.
* Westermann et al. [15] took a different approach to the problem by examining the effect of system-generated reports of connectedness on credibility. The researchers took an experimental approach to designing six mock-up pages on Twitter that varied the ratio between followers and follows and the number of followers. The results revealed that having too many followers or too few led to low assessments of expertise and trustworthiness. Having a narrow gap between follows and followers led to higher assessments of credibility.
* Kang et al. [16] discussed ways to model topic-specific credibility on Twitter on an evaluation of three computational models such as a social model, a content-based model, and a hybrid model. The authors used seven-topic specific data sets from
* Twitter to evaluate these models. The results showed that the social model outperformed the others in terms of predictive accuracy.
* Ikegami et al. [17] performed a topic- and opinion classification-based credibility analysis of Twitter tweets, using the Great Eastern Japan earthquake as a case study. The researchers assessed credibility by computing the ratios of similar opinions to all opinions on a particular topic. The topics were identified using latent Dirichlet allocation (LDA). Sentiment analysis was performed using a semantic orientation dictionary to assess whether a tweet’s opinion was negative or positive. An evaluation of this method using kappa statistics showed that it is a good way to assess credibility.

**Disadvantages**

* + There is no Credibility and Reputation.
  + There is no Sentiment Analysis.

**PROPOSED SYSTEM**

* In the proposed system, the system proposes a new credibility analysis system for assessing information credibility on Twitter to prevent the proliferation of fake or malicious information. The proposed system consists of four integrated components: a reputation-based component, a credibility classifier engine, a user experience component, and a feature-ranking algorithm.
* The components operate together in an algorithmic form to analyze and assess the credibility of Twitter tweets and users. We tested the performance of our system on two different datasets from 489,330 unique Twitter accounts. We applied 10-fold cross-validation over four machine learning algorithms. The results reveal that a significant balance between recall and precision was achieved for the tested dataset.
* The system also proposes a novel credibility assessment system that maintains complete entity-awareness (tweet, user) in reaching a precise information credibility judgment. This model comprises four integrated components, namely, a reputation- based model, a feature ranking algorithm, a credibility assessment classifiers engine, and a user expertise mod-el. All of these components operate in an algorithmic form to analyze and assess the credibility of the tweets on Twitter.

**Advantages**

* The system validated by applying tenfold cross validation

with machine-learning algorithms..

* Credibility of the source during an event.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**Software Requirements:**

* Operating System - Windows XP
* Coding Language - Java/J2EE(JSP,Servlet)
* Front End - J2EE
* Back End - MySQL