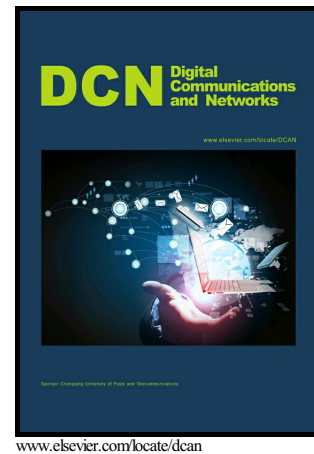


Author's Accepted Manuscript

Sarcastic Sentiment Detection in Tweets Streamed
in Real time: A Big Data Approach

S.K. Bharti, B. Vachha, R.K. Pradhan, K.S. Babu,
S.K. Jena



PII: S2352-8648(16)30027-X
DOI: <http://dx.doi.org/10.1016/j.dcan.2016.06.002>
Reference: DCAN42

To appear in: *Digital Communications and Networks*

Received date: 20 February 2016
Revised date: 16 May 2016
Accepted date: 15 June 2016

Cite this article as: S.K. Bharti, B. Vachha, R.K. Pradhan, K.S. Babu and S.K. Jena, Sarcastic Sentiment Detection in Tweets Streamed in Real time: A Big Data Approach, *Digital Communications and Networks* <http://dx.doi.org/10.1016/j.dcan.2016.06.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain



Sarcastic Sentiment Detection in Tweets Streamed in Real Time: A Big Data Approach

S K Bharti * ^a, B Vachha^a, R K Pradhan^a, K S Babu^a, S K Jena^a

^aDepartment of Computer Science & Engineering, National Institute of Technology, Rourkela 769008, India

Abstract

Sarcasm is a type of sentiment where people express their negative feelings using positive or intensified positive words in the text. While speaking, people often use heavy tonal stress and certain gestural clues like rolling of the eyes, hand movement, etc. to represent sarcastic. In the textual data, these tonal and gestural clues are missing, making sarcasm detection very difficult for an average human. Due to these challenges, researchers are showing interest in sarcasm detection of social media text, especially in tweets. Rapid growth of tweets in volume and its analysis pose major challenges. In this paper, we proposed a Hadoop based framework that captures real time tweets and process it with a set of algorithms which identifies sarcastic sentiment efficiently. We observe that the elapse time for analysing and processing under Hadoop based framework significantly outperforms the conventional methods and is more suited for real time streaming tweets.

© 2015 Published by Elsevier Ltd.

KEYWORDS: Big Data, Flume, Hadoop, Hive, MapReduce, Sarcasm, Sentiment, Tweets

1. Introduction

With the advent of smart mobile devices and the high-speed Internet, users are able to engage with social media services like Facebook, Twitter, Instagram,

*Santosh Kumar Bharti (Corresponding author) is currently pursuing his PhD in Computer Science & Engineering from National Institute of Technology Rourkela, India. His research interest includes opinion mining and sarcasm sentiment detection (email:sbharti1984@gmail.com).

¹Bakhtyar Vachha is currently pursuing his M.Tech in Computer Science & Engineering from National Institute of Technology Rourkela, India. His research interest includes network security and big data. (email:bakhtyarvachha@gmail.com).

²Ramkrushna Pradhan is currently pursuing his M.Tech dual degree in Computer Science & Engineering from National Institute of Technology Rourkela, India. His research interest includes speech translation, social media analysis and big data. (email:ramkrushna768@gmail.com).

³Korra Sathya Babu is working as Asst. Professor in the department of Computer Science & Engineering, National Institute of Technology Rourkela, India. (email:ksathyababu@nitrrkl.ac.in).

⁴Sanjay Kumar Jena is working as Professor in the department of Computer Science & Engineering, National Institute of Technology Rourkela, India. (email:skjena@nitrrkl.ac.in).

etc. The volume of social data being generated is growing rapidly. Statistics from Global WebIndex shows a 17 % yearly increase in mobile users with the total number of unique mobile users reaching 3.7 billion people [1]. Social networking websites have become a well-established platform for users to express their feelings and opinions on various topics such as events, individuals or products. Social media channels have become a popular medium to discuss ideas and to interact with people worldwide. For instance, Facebook claims to have 1.59 billion monthly active users, each one being a friend with 130 people on average [2]. Similarly, Twitter claims to have more than 500 million users, out of which more than 332 million are active [1]. Users post more than 340 million tweets and 1.6 billion search queries every day [1].

With such large volumes of data being generated, a number of challenges are posed. Some of them are accessing, storing, processing, verification of data sources, dealing with misinformation and fusing various types of data [3]. However, almost 80% of gener-

Download English Version:

<https://daneshyari.com/en/article/5000819>

Download Persian Version:

<https://daneshyari.com/article/5000819>

[Daneshyari.com](https://daneshyari.com)