Advanced Science and Technology Letters
Vol.136 (ITCS 2016), pp.41-44
http://dx.doi.org/10.14257/astl.2016.136.11

Analysis on big data by performance factors of creative education using semi-structured data-based Twitter (with focus on Republic of Korea)

Ji-Hoon Seo[1] , Kil-Hong Joo[2] , Nam Hun Park[3]

Dept. of Computer Science Engineering Incheon University, Korea
Dept. of Computer Education, Gyeongin National University of Education, Korea
Dept. of Computer Science, Anyang University, Korea
sserz@inu.ac.kr, nmhnpark@anyang.ac.kr, khjoo@ginue.ac.kr

Abstract. As various forms of big data, which includes but not limited to, large volume texts, voice data and videos, are being accumulated whilst the waves of the information age are accelerating progressively, the number of inter-disciplinary analysis solutions with capabilities to use such information is increasing, and accordingly, the developments, such as the drop of costs required for data storage and various Social Network Services, have brought forth the quantitate and qualitative stretch of the data. The phenomenon makes it possible to achieve the types of data usage which were not available in the past, and thus the potential values and leverage of data are on the rise. Studies that that apply such inter-disciplinary analysis system for the improvement of the educational system to suggest future-oriented education system are being carried out at progressive levels. This study has carried out an analysis on big data with Twitter as its subject and suggested, via the natural language process of data and frequency analysis, the quantitative scale indicative of how various issues and performances relating to creative education in South Korea have been handled.

Keywords: Creative education, big data, pattern algorithm, association role

1 Introduction

As most public institutions of South Korea and its departments have initiated to provide and be equipped with effective systems for information management & analysis and information use in agreement with government 3.0, higher demands for systematic analyses and applications of big data are being made to seek education performance analyses and improvement proposals. The phenomenon not only enables us to make use of the data, which was not attempted previously, and is enlarging the potential values and leverage of the data. There exist many fields in which these aspects can be applied, and measures apt for carrying out appropriate education can be suggested, if the educational data previously accumulated are properly utilized. Methods that have previously existed were consisted of brainstorming, Delphi, panel of experts and et cetera, whereas, in this study, results of analysis that are based on

ISSN: 2287-1233 ASTL

quantitative data can be suggested by analyzing the big data objectively, for it uses Twitter via online mediums[1][2]. Therefore, this study, with creative education, one of the country's policies that is being enforced and highlighted, as its keyword, has made discoveries and carried out extensive researches on issues and performances creative education have achieved, if any, by collecting the South Korean Twitter data consisted of data of the past to the present time.

2 Related Work

Studies that cover big data of educational information include researches carried out by Jaeyeong Choi (2012), Yongsang Jo et al. (2013) and Wooju Kim (2013). Jaeyeong Choi (2012), in his study, has discussed big data in the Smart Education environment. The study estimates the types and sizes of educational data which can occur within the scope of the Smart Education environment and suggests methods for utilizing big data in the Smart Education environment. By having estimated the educational big data that occurs within the scope of the Smart Education environment, it has estimated that seven million elementary, middle and high school students throughout the county produce 7.0 terabytes of data a day and the sum data per year amounts to 2.56 petabytes. In addition, the study has suggested customized learning support, contents development, improvement of study courses and educational policies, and student guidance as methods for utilizing the educational big data [3][4].

3 Proposed Method

In this thesis, a five-phase process, which includes the collection of semi-structured data, securement and classification of historic data through collection, data preprocessing, natural language processing of the Korean grammar and big data analysis, had been carried out via Twitter as a process for the analyzation of performance factors.

3.1 Collecting semi-structured data for big data analysis

Collecting historic data is vital for carrying out a quantitative analysis on the recent analysis of creative education of South Korea. Historic data may have subtle effects on the results of big data analysis based on the amount, cycle and route pursuant to the collection, and the errors of the data being resulted can be applied in a significant way.

3.2 Preprocessing procedures for data collection

The historic data gathered by media performs the preprocessing process, and morpheme analysis and natural language process methods by decomposition in text

mining are required to analyze the semi-structured data on text documents. Unnecessary attribute factors of the data field are removed from the Korean sentences in the preprocessing procedures.

3.3 Preprocessing of natural language in Korean texts

The process for the preprocessing of natural language is carried out on Twitter texts by using the Korean morphological analyzer. The sentences in Twitter texts are mixed with parts of speech, such as subjects, nouns, verbs, adjectives, adverbs, exclamations, etc., and such characteristics are not easy to analyze with the attributes of natural language.

3.4 The execution of collective examination of time series′ recurrent entries

Frequent patterns are patterns that occur frequently in data sets. In other words, by analyzing the terminologies and sentences that appear most frequently in Twitter by having them applied to Twitter, most frequently used frequent keywords regarding creative education are extracted. In addition, by reflecting the fact that they are log-like records of short sentences, sentences and keywords in a long-range time period due to the attributes of Twitter, extraction of Twitter keywords and frequency time series analysis are carried out simultaneously for analyzation.

4 Performance Evaluation

Following is the result of the analyzation of keywords of ‘creative education’. With regard to the time period, twits that were made between year 2014 and 2015 were collected.

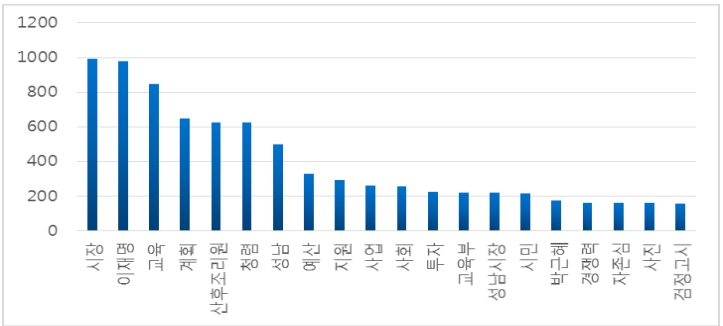

Figure 1. 2014 creative education keyword Top 20

As a result of the analysis made, it has been confirmed that concerns regarding private education expenses and budget on national policies and elementary school

students were highlighted as a social issue in South Korea, and, regardless of the fact that notable achievements regarding'creative education' have not yet been made, it has been analyzed that various areas of education, including but not limited to arts, music and physical education, are conducting the creative education.

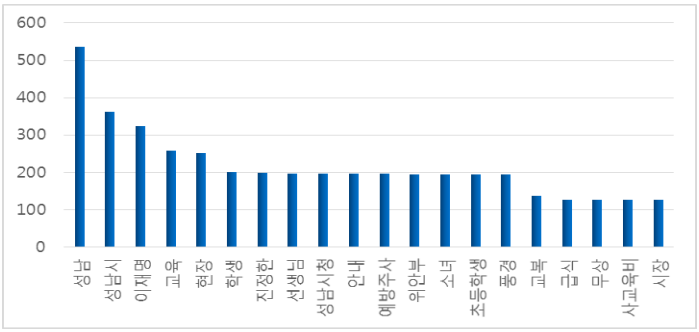

Figure 2. 2015 creative education keyword Top 20

5 Conclusion

This thesis, on the basis of semi-structured data of Twitter comprised of 140 words, had collected the twit data ranging from year 2013 to 2015 and drawn a conclusion, through big data analyzation, on the measurement and performance factors on the subject of to what degree had creative education permeated the Korean educational system. Despite the fact that the so-called creative education is not yet prevalent in the Korean educational system in a broad perspective, it has been confirmed that it is exercising a huge influence in the Korean society in general via fragments of their national policies, software education and et cetera.

References

1. Sun, Y., Jia, K.: 2009. Research of word sense disambiguation based on mining association rules, In: Third International Symposium on Intelligent Information Technology Application workshops, November 21-22, NanChang, China, pp. 86-88.
2. Dunham, M.H.: Data Mining: Introductory and Advanced Topics, Pearson Education Inc. (2003).
3. Tang, C., Liu, C.: Method of Chinese grammar rules automatically access based on association rules, In: Proceedings of the. Computer Science and Computational Technology volume, 1 pp. 265-268 (ISCSCT, Shanghai, Dec. 20-22, 2008), (2008).
4. Xu, Y., Li, Y., Shaw, G.: Reliable representations for association rules. Data & Knowledge Engineering, Volume 70 Issue 6, pp. 555-575. June, 2011.

