USE OF BIG DATA IN EDUCATION EFFICIENCY ANALYSIS

György Molnár, Dávid Sik, Zoltán Szűts, Budapest University of Technology and Economics, Hungary

Structure of the paper

This paper starts by painting a global picture of social media use among students and presents some finding of a survey in order to show how commonly is the topic of teaching and learning represented in discussions. Secondly, it presents the Big Data paradigm. Thirdly, the paper offers a way of determining the sentiments about educational applications using Big Data analysis. Finally, it brings some conclusions.

Introduction

We live in an information society where more and more interaction takes place in network environment. Our learning activities become digital and e-learning is not a novelty anymore. Several elements of our lives are stored in clouds and analyzed by computers and software. Non-formal and informal education is becoming the new paradigm and due to the nature of the internet and e-learning, more and more people can access knowledge. Voices and opinions which have been silent until now have grown louder. Open discussions have been initiated on topics of learning, teaching and education in general. As the internet and Web 2.0 services in particular are the most democratic platforms where arguments can be made that it is no wonder that many of the students state their opinions online in forums about their instructors, teachers, universities, about the curriculum and lectures. Some of them use social media to express their feelings about education, about the examination systems of learning management systems (LMS). The same time educational platforms become more and more open. MOOC courses are offered for free all over the world, where students get knowledge, but the same time give important insights and the data collected this way is really precious to determine whether a course is successful or not.

Using social media is nowadays an integral part of our information society, especially in case of those members who belong to Z generation that feels free to express their opinions or to become a prosumer (producer + consumer). It is typical for this generation to be always online and immediate share a huge amount of information that can be later analyzed.

Everything that is born or produced among the boundaries of the network becomes searchable and can be analyzed. Information penetrates all levels of society. Businesses, politics, government and education migrates to the network. What was previously private (for example opinions and confessions) nowadays is public knowledge and awaits to be

commented and rated (Csepeli, 2015; pp.172-173). It also must be stated that internet does not forget, with some exaggeration one could say that nothing can be deleted and everything can be searched, found and analyzed. Every interaction online leaves a trail, data and can be researched (Dessewffy & Láng, 2015; p.160).

We have been gaining insights what students think about some elements of education for some time. Students are regularly surveyed in order to determine their opinion about and satisfaction with their professors and classes. But usually this results are biased, or *forced out*. Many of the students are afraid to be completely honest or don't feel the urge to express their opinion. But among the boundaries of social media where they feel at home, safe and free to talk they provide useful insights. These insights are often shattered on million Twitter channels and Facebook pages, blogs and forum, but still they provide knowledge about specific lectures, professors and institutions, and if analyzed though the lenses of Big Data a global picture about for example the usefulness of Moodle (LMS) or Coursera (MOOC) can be painted.

Opinions about learning posted in social media

In order to prove that majority of students express their opinions regarding learning in social media a research has been conducted by electronic survey in 2016 spring. It was based on simple random sampling; the target group involved full time and part time students studying three majors at Budapest University of Technology and Economics: pedagogy, economics and engineering. The research mostly focused on social media using attitudes. The survey was carried out to support our hypothesis that students nowadays argue about educational issues in social media and these opinions can be classified and analyzed. N = 119 analyzable answers arrived within the deadline. The survey consisted of 16 closed questions. The main results regarding this paper are as follows.
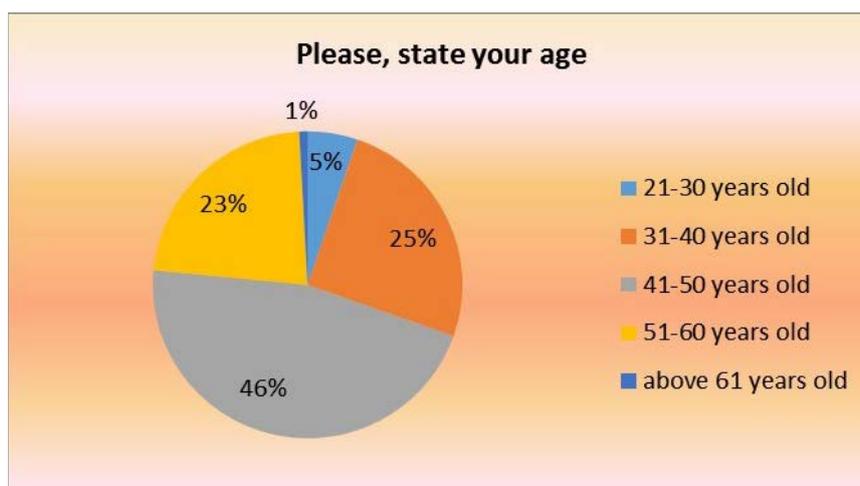

Figure 1. The diagram above shows the age of the respondents.

The majority of the students stated that community media provides a suitable field for retrieving information about courses, topic and professor and provides a democratic platform for discussing various opinions in connection with learning.
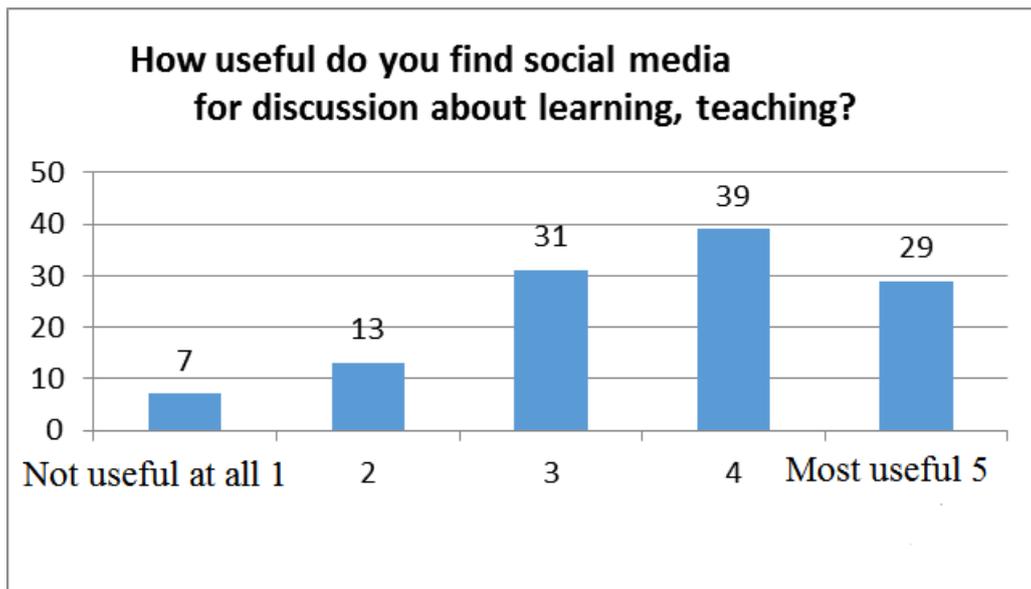
Figure 2. Social media for discussion about learning, teaching

60% of the respondents confirmed daily use of social media. In information society that is based on u-computing this activity is supported by mobile tools such as smart phones or tablets.

Big Data

Our survey was based on the answers of Hungarian students, but it is safe to say that use of social media and the attitude which considers students expressing their opinions about education freely is a worldwide phenomenon. This fact means that huge amount of date on the topic is generated continuously. Social media and the wide variety of social media channels exist where a huge amount of data available. The challenge comes in accessing that data and transforming it into something that is usable and actionable (Thiel et al., 2012). In order to analyze such a big amount of information, we should turn to Big Data.

Big Data is nowadays a popular term that describes the growth, availability and analysis of both structured and unstructured data. A huge amount of data (information) is being collected continuously from several sources, for example web-browsing and searching, social media, banking, air traffic. This data is than stored and evaluated. The amount of data is intangible in terms of personal computing. A look at the numbers reveals its potential and the amount of work needed in order to classify data and get valuable insights.

2.5 quintillion bytes of data are created every day. The growth of information becomes clear if one points out that 90% of the data in the world today has been created in the last 2 years alone. Data comes from several sources: machines and users. IoT (Internet of Things), bank transfers as well as social media interaction creates Big Data (IBM, n.d.). This paper considers the latest to be useful in order to analyze the opinions of users regarding education. Data created by users in social media can be used to determine sentiments. Big Data tools – some open source – are now available to gain a first impression of a particular social media channel.

This way opinions posted on Twitter or Facebook, Google can be studied to provide an overview (Thiel et al., 2012).

Big Data is more than simply a matter of size; it is an opportunity to find insights in new and emerging types of data and content, to make businesses more agile, and to answer questions that were previously considered beyond our reach. Until now, there was no easy way to harvest this knowledge. But today we are witnessing an exponential growth in the volume and details of data captured by social media (Bessis & Dobre, 2014; p.4).

Usually, Big Data is characterized by three main properties, referred to as the three Vs (Laney, 2001):

- Volume (data size or amount of data);
- Velocity (the fast rate at which data are generated or they need to be processed);
- Variety (heterogeneity in content and/or form).

Data today comes in all types of formats – from traditional databases to hierarchical data stores created by end users such as camera feeds, news or social media posts – in our paper we analyze the latest.

There is a 4th V: Variability. In addition to the increasing velocities and varieties of data, data-flow can be highly inconsistent with periodic peaks. In case of social media this varies daily and event triggered peak data loads are usual (Majkic, 2014).

Social media opinions analysis using Big Data

Collecting and analyzing data stream of social media holds the promise of gaining more, faster and better insights. With the Twitter platform and blogs especially, much of this information is freely available, in case of Facebook it is more complicated as its algorithm in not really open for everyone to see. Twitter allows researchers to explore novel means of analyzing media content, as they use computational methods to assemble, filter, and interpret much of the collective Twitter conversation around a particular topic or event (Lewis et al., 2013).

Efficiency of educational methods – sentiment analysis

Opinion mining or sentiment analysis is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities such as products, services, organizations, individuals, events, and their different aspects. It has been an active research area in natural language processing and Web mining in recent years. However, people have difficulty, owing to their mental and physical limitations, producing consistent results when the amount of such information to be processed is large. Automated opinion mining is thus needed, as subjective biases and mental limitations can be overcome with an objective opinion mining system (Zhang & Liu, 2014; p.1).

The Big Data sentiment analysis presented in this paper is focused on teaching-learning issues. Using *sentiment viz* (https://www.csc.ncsu.edu/faculty/healey/tweet_viz/tweet_app/) tool we searched for general opinion about Moodle and midterms in general.

Figure 3. Sentiments about Moodle

The search for opinions about Moodle was really productive. The search returned 277 almost real time (within a time span of an hour) hits where users tweeted using the word Moodle, with the majority of tweets them being associated with positive feelings and shoving mostly active involvement. Most of the students are on alert or exited while using Moodle, they feel calm or relaxed. The most negative opinions are related to the state of boredom and intensity. Not one user expressed a feeling where he was nervous or depressed.
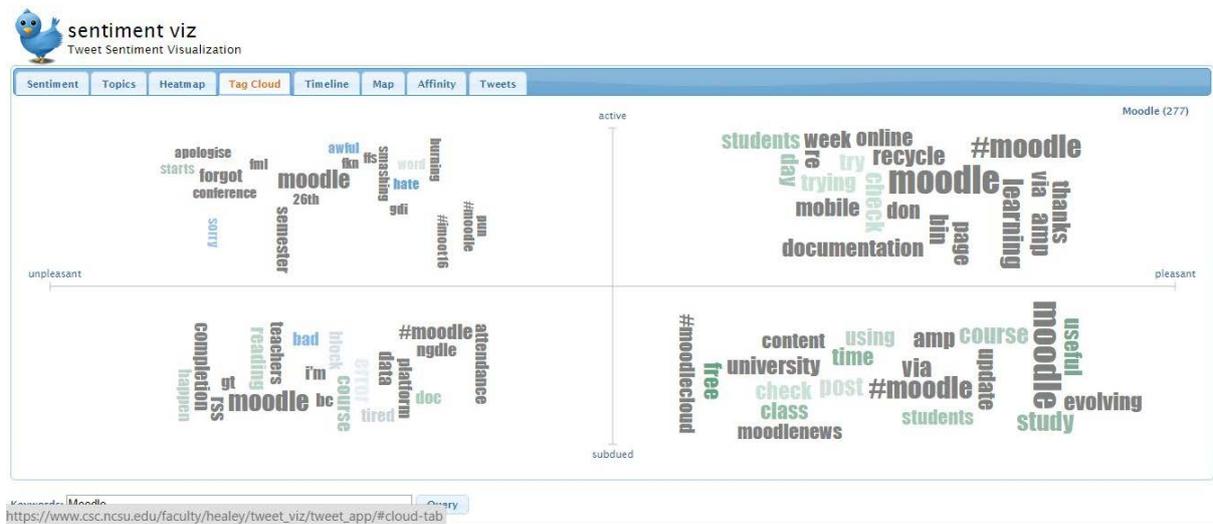

Figure 4. Tag clouds associated with the word Moodle

The tweets characterised as pleasant and active centre around tags such as learning, mobile, documentation, trying or check. The one which are pleasant but subdued use words useful, evolving, free, content, time or class. There were just a few negative tweets. The ones that carried active sentiments used word such as: awful, smashing, burning or forgot. Finally, the negative and subdued one mentioned Moodle together with the terms: teachers, course, attendance or reading. According to opinions, Moodle is considered to be a pleasant way of learning, where users are even exited, and no one is stressed or unhappy.

Figure 5. Sentiments about grammar

In case of sentiments about grammar in general, the results show a bit different picture than in case of Moodle. Even though the feelings expressed are mostly positive and some tweeters admitting to be elated or serene, there are some negative clusters of information, and many users feel tense or bored when expressing their opinions about grammar. Words that are associated with the state where the user feels depressed or nervous are also used.
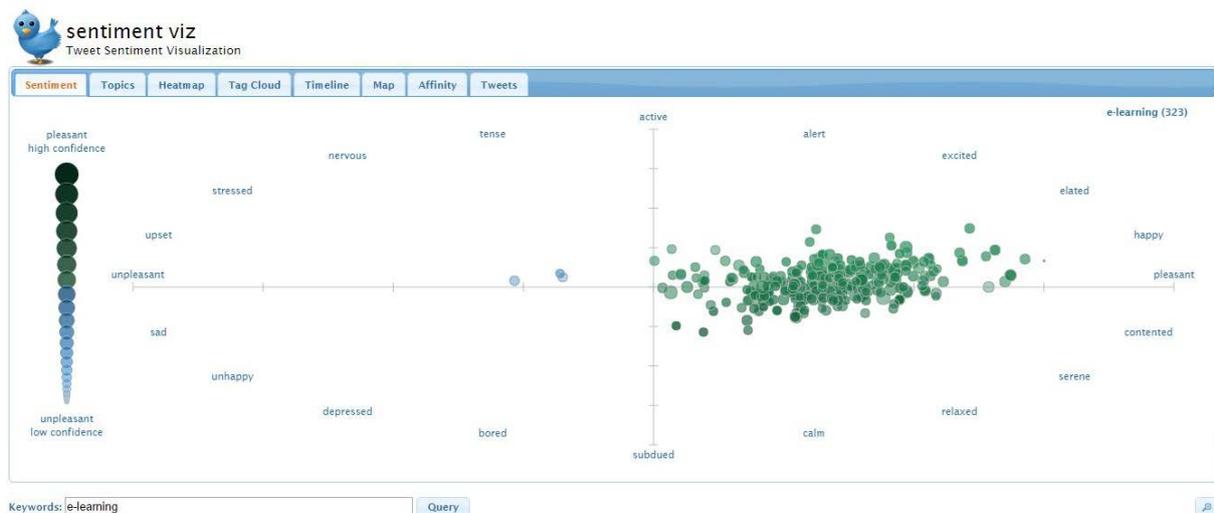
Figure 6. Sentiments about e-learning

The feelings expressed while tweeting about e-learning are the most homogenous. Users feel alerted, calm or exited. And only 3 tweeters expressed to be tense or bored. No one stated to be nervous or depressed. Results gained through Big Data analysis show that e-learning is considered to be a very posity element of education.

Conclusions

Examining the results we can sum up that there is an intensive chatter on Twitter about Moodle, grammar or e-learning in general. One would assume that they would not be popular topics, but with 7,000 tweets send in one second there is an amount of tweets about them that

can be analyzed. Using Big Data helps gather the information and then analyze is to get valuable insights. As several applications are being develop to handle Big Data, such as Hadoop and machine learning becomes more efficient, a clearer picture can be seen from information shattered all over social media. Insights got this way are precious as they can be used both by teachers and students.

References

1. Benedek, A., Molnár, G., & Szűts, Z. (2015). Practices of Crowdsourcing in relation to Big Data Analysis and Education Methods, In A. Szakál (Ed.), *SISY 2015: IEEE 13th International Symposium on Intelligent Systems and Informatics: Proceedings* (pp. 167-172). Subotica: IEEE Hungary Section, 2015.

2. Berro, A., Megdiche, I., & Teste, O (2015). A content-driven ETL processes for open data. In N. Bassiliades, M. Ivanovic, M. Kon-Popovska, Y. Manolopoulos, T. Palpanas, G. Trajcevski, & A. Vakali (Eds.), *New Trends in Database and Information Systems II, AISC, vol. 312* (pp. 29-40). Heidelberg: Springer.

3. Bessis, N., & Dobre, C. (2014). Big Data and Internet of Things: A Roadmap for Smart Environments. Springe.

4. Ciobanu, R-I., Cristea, V., Dobre, C., & PopBessis, F. (2014). Big Data Platforms for the Internet of Things. In N. Bessis & C. Dobre (Eds.), *Big Data and Internet of Things: A Roadmap for Smart Environment* (pp. 3-34). New York, Springer.

5. O'Leary, D. E. (2013). 'Big data', the 'internet of things' and the 'internet of signs'. *Intelligent Systems in Accounting, Finance and Management, 20*(1), 53-65.

6. IBM (n.d.). *What is big data?* Retrieved from http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

7. Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety.* Meta Group. Retrieved from https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

8. Lengyel, L., Ekler, P., Ujj, T., Balogh, T., Charaf, H., Szalay, Z., & Jereb, L. (2015). ICT IN ROAD VEHICLES – The VehicleI ICT Platform. *Proceedings of the Models and Technologies for Intelligent Transportation Systems (MT-ITS) 3-5, June 2015, Budapest, Hungary*, 457-462.

9. Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods. *Journal of Broadcasting & Electronic Media, 57*(1), 34-52.

10. Majkić, Z. (2014). Big Data Integration Theory and Methods of Database Mappings, Programming Languages, and Semantics. New York: Springer.

11. Szűts, Z., & Molnár, G. (2016). Hatékony tanulási és tanítási módszerek vizsgálata a közösségi média és Big Data környezetében, In L. Hülber & A. Tamásné Fekete (Eds.), *I. Oktatástervezési és Oktatás-informatikai Konferencia: Absztraktkötet.*

12. Thiel, K., Kötter, T., Berthold, M., Silipo, R., & Winters, P. (2012). Creating Usable Customer Intelligence from Social Media Data: Network Analytics meets Text Mining. KNIME. Retrieved from https://www.knime.org/files/knime_social_media_white_paper.pdf

13. Ujbanyi, T., Katona, J., Kővári, A., Király, Z., & Kadocsa, l. (2014). IKT-eszközök bevezetésének és használatának problémái az oktatásban, In N. Kiss, B. Nagy, & I.P. Németh (Eds.), *Tudományos terek* (pp. 21-34). Dunaújváros: DUF Press. Dunakavics könyvek, 6.

14. Westera, W. (2013, October 14). Social Media and Big Data – Cracks in the Crystal Ball? [Blog post] RW Connect. Retrieved from https://rwconnect.esomar.org/using-social-media-for-market-analysis-cracks-in-the-alleged-crystal-ball/

15. Zhang, L., & Liu, B. (2014). Aspect and Entity Extraction for Opinion Mining. In W. W. Chu (Ed.), *Data Mining and Knowledge Discovery for Big Data – Methodologies, Challenge and Opportunities.* New York: Springer.