Predicting Student Performance and Differences in Learning Styles based on Textual Complexity Indices applied on Blog and Microblog Posts

A Preliminary Study

Elvira Popescu[1], Mihai Dascalu[2], Alex Becheru[1], Scott Crossley[3], Stefan Trausan-Matu[2]
[1] University of Craiova, Romania, popescu_elvira@software.ucv.ro, becheru@gmail.com
[2] University Politehnica of Bucharest, Romania, mihai.dascalu@cs.pub.ro, stefan.trausan@cs.pub.ro
[3] Georgia State University, Atlanta, USA, scrossley@gsu.edu

***Abstract*—Social media tools are increasingly popular in Computer Supported Collaborative Learning and the analysis of students' contributions on these tools is an emerging research direction. Previous studies have mainly focused on examining quantitative behavior indicators on social media tools. In contrast, the approach proposed in this paper relies on the actual content analysis of each student's contributions in a learning environment. More specifically, in this study, textual complexity analysis is applied to investigate how student's writing style on social media tools can be used to predict their academic performance and their learning style. Multiple textual complexity indices are used for analyzing the blog and microblog posts of 27 students engaged in a project-based learning activity. The preliminary results of this pilot study are encouraging, with several indexes predictive of student grades and/or learning styles.**

Keywords - social media; textual complexity analysis; student performance; learning style

I. Introduction

Social media tools are increasingly being used in collaborative learning environments due to their support for student-generated content and their potential for enhancing communication, sharing, and cooperation. According to Dron and Anderson [4], the main pedagogical contributions of social media for learning include creating and shaping communities, as well as building knowledge. Furthermore, social media can foster positive interactions between learners and increase learner engagement and motivation. Social media also encourages active learning by facilitating activities such as debates, problem-solving tasks, and/or inquiries [4]. Various collaborative learning scenarios have been designed around social media tools such as wikis [26], blogs [24], social bookmarking services [5] or microblogging services [10].

With the expanding popularity of social media tools in Computer Supported Collaborative Learning (CSCL), the need to examine how individual learners interact differently in online communities has increased. In this context, the aim of this study is to evaluate how the learners' writing style in social media environments can be used to predict their overall learning performance, as well as their individual learning style. To do so, we use multiple textual complexity indices ranging from lexical, syntactical to semantic analyses [16], described in detail later on in this paper, to create an in-depth perspective of each learner's writing style. This style analysis creates a basis for predicting the overall performance of each learner in their educational scenario, as well as for highlighting specific traits of their learning style.

According to Keefe [11], learning styles refer to the individual manner in which a person approaches a learning task, as well as their preferences related to perception modality, processing and organizing information, reasoning or social aspects. Various learning style models have been proposed during the past decades, which differ in terms of the underlying learning theories, as well as the number and descriptions of included dimensions. One of the most popular models in technology-enhanced learning is FSLSM proposed by Felder and Silverman [3, 6]. According to FSLSM, learners are categorized in terms of their preferences based on four dimensions: a) *active* versus *reflective*; b) *sensing* versus *intuitive*; c) *visual* versus *verbal* and d) *sequential* versus *global* [6].

So far, few studies have explored the relations between students' behavior in social media-based learning environments (i.e., patterns of interaction with the tools) and their learning styles [3, 20]. Previous studies have found that: i) *active* students tend to post more frequently to their blogs than *reflective* students; ii) *reflective* students' ratio of reading other blog postings vs. posting to their own blogs is significantly higher than that of *active* students; iii) *active* students use charts displaying the number of postings and peer rating more often than *reflective* students; iv) *sequential* learners tend to write longer posts than *global* learners [3].

Similarly, only a few studies have explored the relationship between students' active participation on social media tools and their academic performance [9, 10]. Preliminary results suggest that the number of blog posts, wiki page revisions, and shared bookmarks are reliable predictors of student success [9]. In terms of analysis techniques, previous studies have relied on rank correlation analysis [3], mixed effects analysis of variance (ANOVA) models [10], principal component analysis [9], as well as machine learning algorithms for classification, association rule induction and feature selection [20].

However, examining quantitative behavior indicators is only one approach to assessing learning behavior and success. The approach that we propose in this paper is based

2161-377X/16 $31.00 © 2016 IEEE
DOI 10.1109/ICALT.2016.104

on the content analysis of each student's contributions as opposed to student behaviors. Specifically, in this study, we apply textual complexity analyses on blog entries and tweets posted by 27 students in the context of a collaborative project-based learning scenario. We also use survey results to assess students' learning styles. We use the textual analysis to predict learning outcomes and learning styles.

The rest of the paper is structured as follows. An overview on textual complexity evaluation is included in section 2. The context of study, together with the data collection and preprocessing steps are described in section 3. The results of the analysis are reported and discussed in section 4. Finally, section 5 concludes the paper and outlines future research directions.

II. Textual Complexity Assessment

Textual complexity analysis can be used both for identifying the most appropriate reading material according to students' comprehension level and for assessing students' writing style and knowledge level from their writing traces, which is our current research goal. According to [16], measuring textual complexity can be split into three perspectives: qualitative, reader/task orientation, and quantitative. The qualitative dimensions of textual complexity cover various levels of meaning, structure, language conventionality and knowledge requirements. Reader and task considerations are associated with readers' motivation, knowledge and interest. Quantitative factors are the core of our analysis because they create the basis for automated methods and tools which can be used to analyze textual complexity.

Such automated methods have been developed into frameworks that have demonstrated success in understanding learning behaviors in collaborative learning environments [17]. These frameworks include: *Lexile* (MetaMetrics), *ATOS* (Renaissance Learning), *Degrees of Reading Power: REAP* (Carnegie Mellon University), *SourceRater* (Educational Testing Service), *Coh-Metrix* (University of Memphis), *TAALES* and *TAACO* (Georgia State University). The implemented framework that we use in this study, *ReaderBench* [2], covers a wide range of lexical, syntactic, semantics and discourse centered textual complexity indices, including the most frequent indices from the above-mentioned systems. The following subsections present the integrated dimensions of analyses possible within *ReaderBench* (RB).

A. Surface Analysis

Categories like fluency, diction and basic readability formulas relying on surface indices (e.g., words, commas, phrase length, periods) are computed in RB as a way of evaluating lexical and syntactic levels of text difficulty. Page and Wresch [18, 28] have demonstrated that static attributes can effectively predict essay scores. Page's work on quantifying an essay's complexity has led to the identification of correlations between *proxes* (computer approximations of interest) and *human trins* (intrinsic variables – human measures used for evaluation). Starting from Page's metrics and taking into consideration Slotnick's categories [22] of grouping proxes based on their intrinsic values, multiple indices from their studies have been integrated within our model including: average paragraph/ sentence/word lengths in characters, average and standard deviation of paragraph/sentence lengths in words (including separate indices for unique content words), as well as number of commas per sentence/paragraph. In addition, entropy [21], which can be defined as the expected value of the information contained in the text, is used in RB to evaluate lexical diversity of word stems and characters within the input text.

B. Word Complexity

In the first step of the RB processing, words are extracted from the input text. A Natural Language Processing (NLP) pipeline then splits the words, tokenizes them, eliminates stop words, and then conducts Part Of Speech (POS) tagging, lemmatization, parsing and Named Entity Recognition (NER) [13]. RB calculates a number of indices for content words in the text. These indices include *syllable count, distance between the inflected form, lemma* and *stem, corpus frequency, distance within the hypernym tree* and *word polysemy*. These indices are used to approximate each word's individual complexity. *Corpus frequency* is computed as an inverse frequency of words from the training corpora, while *polysemy* is computed on the basis of the lexicalized ontology WordNet [15]. In general, words with multiple senses have a higher complexity because it is harder to assign the correct sense to them. The *distance within the hypernym tree* is related to a word's specificity and is determined as the distance to the ontology's root; longer paths usually indicate specialization or specificity for given words. In addition, the differences between the *inflected form,* the *lemma* and the *stem* reveal the use of multiple juxtaposed prefixes and suffixes, which is another mark of a word with a higher complexity.

C. Syntax

The most predictive POS tags in terms of textual complexity are prepositions, adjectives and adverbs [17] that allow for a more detailed and complex text structure. In addition, RB uses a syntactic parser to calculate a number of syntactic indices (e.g., *overall size of the parsing tree, maximum depth, number of semantic dependencies).* Higher values for these indices usually indicate greater complexity [8].

D. Semantics

According to McNamara et al. [14], *textual complexity* is also linked to *cohesion* with regard to text comprehension. Cohesion is a central element for obtaining a coherent mental representation of discourse, commonly called a situation model [27]. Therefore, texts that lack cohesion may be perceived as having a higher difficulty due to an increased cognitive load on the part of the reader. In RB, a cohesion graph is used to model the underlying structure of discourse [25]. Cohesion is determined at both inter- and intra-paragraph levels of analysis based on three semantic distances: a) Wu-Palmer semantic similarity in WordNet

[29], b) cosine similarity using Latent Semantic Analysis vector spaces [12] and c) Jensen-Shannon dissimilarity between Latent Dirichlet topic distributions [1]. In general, a text is more complex as the number of relevant links in the cohesion graph increases. Thus, the average value of all the inter- and intra- paragraph links reflects text complexity [25]. In addition, specific discourse connectors defined as cue phrases are also considered in order to evaluate the degree of discourse elaboration.

Named entity derived features also influence comprehension, because they create the basic components of concepts and propositions on which higher-level discourse processing is based [7]. Therefore indexes based on named entity statistics are also included in the RB framework.

III. Experimental Settings and Data Collection

A. Context of Study

Data for this study was collected during the first semester of the 2013-2014 academic year at the University of Craiova, Romania. The educational context was a course on "Web Applications Design" (WAD) taught to 4th year undergraduate students in Computer Science. A project-based learning (PBL) approach was used in which learning was organized around the development of web applications comprising of multiple state-of-the-art technologies. Students collaborated in teams of 4 peers in order to build their chosen application (e.g., a virtual bookstore, an online auction website, a professional social network, an online travel agency, etc.).

The PBL scenario was implemented in blended mode, with weekly face-to-face meetings between each team and their instructor. These meetings were complemented by the use of three social media tools (wiki, blog and a microblogging tool) for online communication and collaboration. MediaWiki[1] was used for collaborative writing tasks, for gathering and organizing team knowledge-base and resources, and for documenting the project. Blogger[2] was used for reporting the progress of each project similar to a "learning diary", for publishing ideas and resources, as well as for providing feedback and solutions to peer problems. Each team had its own blog, but inter-team cooperation was encouraged as well. Twitter[3] was meant for encouraging additional connects to peers and for posting short news, announcements, questions, and status updates regarding each project.

These three CSCL tools were all integrated in a social learning environment called eMUSE (**e**mpowering **M**ash**U**ps for **S**ocial **E**-learning) [19]. The platform provided support for both students and teachers as a unique access point to the social media tools, basic administrative services, learner tracking and data visualizations, as well as evaluation and grading support.

In addition, students had to create four compulsory intermediary presentations in order to be actively engaged throughout the semester and to discourage the practice of activity peaks at the end. Each student's performance assessment took into account both the final product delivered at the end of the semester and the continuous collaborative work carried out on the social media tools made available in eMUSE.

B. Data Collection

The participants in the study consisted of 66 undergraduate students split into 17 teams (16 with 4 members and 1 with 2 members), who were enrolled in the WAD course. All student actions on the three social media tools were monitored and recorded in the eMUSE platform. The system gathers learner actions from each of the disparate CSCL tools and stores them in a local database, together with a description and an associated timestamp. For the current study, the writing actions we used to assess student writing style were students' tweets, together with their blog posts and comments. The total number of student contributions recorded at the end of the semester included 1561 tweets, 708 blog posts and 366 blog comments.

To collect students' learning styles according to FSLSM [6] we used a dedicated inventory: the Index of Learning Styles questionnaire (ILS) [23]. The ILS consisted of 44 questions, each with two possible answers. As a result of the test, the learning style of the student is described on a scale between -11 and +11 (with a step of +/-2) for each FSLSM dimension. As an example, a score of +11 on the *active/reflective* dimension implies a strong *active* preference, while a score of -3 implies a mild *reflective* preference. The ILS questionnaire was not mandatory, and, as a result, only 48 students completed it.

C. Data Preprocessing

Data preprocessing was necessary because many texts contained elements that added no value to our research (e.g., HTML tags, images, emoticons, references to other students, computer code, URLs). Also, students used both English (second language) and Romanian (native) languages in their blog and Twitter contributions. However, textual complexity cannot be performed in a cross-language manner due to each language's specificities. Thus, we had to separate English from Romanian texts. For the purpose of this study, only English contributions were considered.

The first preprocessing step focusing on cleaning the input texts was automatically performed by means of an extension of the BeautifulSoup[4] library. After applying the previous cleaning mechanisms, we performed language detection based on the langdetect[5] library (ported from Google's language detection software).

The second step centered on making language corrections (i.e., spelling) and was done manually. Besides spellchecking, we also enriched the texts with expanded Twitter tags, if present. For example, if a semantically relevant Twitter tag was found (e.g., #workinghard), then the words contained within the tag were extracted and added

[1] http://www.mediawiki.org
[2] http://www.blogger.com
[3] https://twitter.com

[4] https://pypi.python.org/pypi/beautifulsoup4
[5] https://pypi.python.org/pypi/langdetect

to the text. Finally, only students who had at least five English contributions after the preprocessing step and used at least 50 content words were considered in order to meet the minimum content threshold needed for our textual complexity analysis. Thus, our statistical analysis described in the following section was performed on 29 students having 848 textual contributions, out of which 27 finished the course and only 25 completed the ILS questionnaire.

IV. Results and Discussion

Due to the limited amount of data and the need to control for over-fitting, 2-tailed Pearson correlations were first computed in order to determine which textual complexity indices were highly correlated to the project and exam grades (see Table I).

TABLE I. Correlations between textual complexity indices and project / exam grades (N = 27)

	Project Grade	*p*	Exam Grade	*p*
Average named entities per contribution	.744**	0	.575**	.002
Average commas per contribution	.517**	.006	.371	.057
Average unique words per contribution	.502**	.008	.413*	.032
Average indefinite pronouns per contribution	.493**	.009	.430*	.025
Average first person pronouns per contribution	.488**	.01	.434*	.024
Average reason and purpose connectives per contribution	.482*	.011	.355	.069
Average sentence – contribution cohesion via LDA	-.481*	.011	-.423*	.028
Average sentences per contribution	.478*	.012	.481*	.011
Average temporal connectors per contribution	.464*	.015	.377	.052
Average coordinating connectives per contribution	.376	.054	.325	.098
Word entropy	.363	.063	.289	.143
Average unique words per sentence	.335	.088	.179	.371
Average simple subordinators per contribution	.333	.089	.289	.144

Two stepwise regression analyses were performed in order to determine the degree to which the three automated indices with the highest correlations predicted students' project and exam grades. Both regressions yielded significant models, $F_{project\ grade}(1, 25) = 30.981$, $p < .01$, $r = .744$, $R^2 = .553$; $F_{exam\ grade}(1, 25) = 12.331$, $p < .01$, $r = .575$, $R^2 = .330$. For each regression, only one variable was a significant predictor (*average named entities per contribution).* This index accounted for 55% of the variance in the project grading, [$\beta = .744$, $t(1, 25) = 5.46$, $p < .01$] and 33% of the exam grade.

Afterwards, a stepwise Discriminant Function Analysis (DFA) was used to classify students on each learning style dimension. For this analysis, +/-1 and +/-3 values (as resulted from the ILS questionnaire) were considered neutral, whereas the other values per dimension were catalogued as positive/negative. As with the regression analysis, only the top 3 indices with the highest effect size were considered in order to control for overfitting.

A. Active/Reflective Dimension

The stepwise DFA retained two variables and removed the remaining variables as non-significant predictors. The results demonstrate that the DFA using these two indices correctly allocated 15 of the 25 texts in the total set for an accuracy of 60% (the chance level for this analysis is 33%). See Table II for variables and results. The measure of agreement between the actual text type and that assigned by the model produced a weighted Cohen's Kappa of 0.256, demonstrating fair agreement.

TABLE II. DFA results (*N* = 25)

FSLSM dimension	Significant predictors	χ^2	*p*	*Accuracy*
active - reflective	Average conjunct connectives per contribution Average conditional connectives per contribution	25.781 (df = 4)	.001	60%
sensing - intuitive	No predictive model could be trained			
visual - verbal	Average conjunct connectives per contribution Average conditional connectives per contribution Average disjunction connectives per contribution	32.469 (df = 4)	.001	80%
sequential - global	Average reason and purpose connectives per contribution	10.164 (df = 2)	.038	48%

B. Sensing/Intuitive Dimension

The stepwise DFA indicated that no variables were predictive of the sensing – intuitive dimension.

C. Visual/Verbal Dimension

The stepwise DFA retained three variables and removed the remaining variables as non-significant predictors. The results demonstrate that the DFA using these three indices correctly allocated 20 of the 25 texts in the total set for an accuracy of 80% (the chance level for this analysis is 33%). See Table II for variables and results. The measure of agreement between the actual text type and that assigned by the model produced a weighted Cohen's Kappa of 0.583, demonstrating moderate agreement.

D. Sequential/Global Dimension

The stepwise DFA retained one variable and removed the remaining variables as non-significant predictors. The results demonstrate that the DFA using this index correctly allocated 12 of the 25 texts in the total set for an accuracy of 48% (the chance level for this analysis is 33%). See Table II for variables and results. The measure of agreement between the actual text type and that assigned by the model produced a weighted Cohen's Kappa of 0.188, demonstrating fair agreement.

V. CONCLUSIONS

This paper investigated how students' writing style on social media tools can be used to predict their academic performance and learning style. Textual complexity analyses were applied on the blog and microblog posts of 27 students engaged in a project-based learning activity. The results are encouraging: several significant correlations were found between textual complexity indices and project / exam grades and one index (*average named entities per contribution*) was determined to be a significant predictor. Similarly, several indexes were predictive for three of the four FSLSM dimensions: two predictors for the *active/reflective* dimension, three predictors for the *visual/verbal* dimension, and one predictor for the *sequential/global* dimension; no variables were predictive of the *sensing/intuitive* dimension. As expected, since the students' online participation was directly linked to the project grading, it was normal that the effect size for scoring projects is greater compared to the evaluation relying on the final exam grades.

It should be noted that the presented experiment was only a pilot study with a small sample size; therefore, its power is low. Nevertheless, the employed mechanisms are extensible and can be easily applied on a larger student population. Moreover, a larger scale experiment is already undergoing and the corresponding student traces will be subject to validation.

ACKNOWLEDGMENT

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-2604 and by the EC H2020 project RAGE (Realising an Applied Gaming Eco-system), http://www.rageproject.eu, grant agreement no. 644187.

REFERENCES

[1] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet allocation", Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003.

[2] M. Dascalu, Analyzing Discourse and Text Complexity for Learning and Collaborating, Studies in Computational Intelligence, vol. 534. Switzerland: Springer, 2014.

[3] M. Derntl and S. Graf, "Impact of learning styles on student blogging behavior", Proc. ICALT 2009, pp. 369-373, IEEE CS Press, 2009.

[4] J. Dron and T. Anderson, Teaching Crowds. Learning and Social Media. AU Press, Athabasca University, 2014.

[5] F. Dujardin, K. Edwards, and S. Beckingham, "Diigo: social bookmarking in a professional communication course", in C. Cheal, J. Coughlin, S. Moore (Eds.): Transformation in Teaching: Social Media Strategies in Higher Education, pp. 243-274, Informing Science Press, 2012.

[6] R.M. Felder and L.K. Silverman, "Learning and teaching styles in engineering education", Engineering Education, vol. 78(7), 1988, preceded by a preface in 2002: www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/LS-1988.pdf.

[7] L. Feng, M. Jansche, M. Huenerfauth, and N. Elhadad, "A comparison of features for automatic readability assessment", in Proc. 23rd Int. Conf. on Computational Linguistics (COLING 2010), pp. 276–284, Beijing, China: ACL, 2010.

[8] V. Gervasi and V. Ambriola, "Quantitative assessment of textual complexity", in M. L. Barbaresi (Ed.): Complexity in Language and Text, pp. 197–228, Pisa, Italy: Plus, 2002.

[9] C. Giovannella, E. Popescu, and F. Scaccia, "A PCA study of student performance indicators in a Web 2.0-based learning environment", in Proc. ICALT 2013, pp. 33-35, 2013.

[10] R. Junco, G. Heibergert, and E. Loken, "The effect of Twitter on college student engagement and grades", Journal of Computer Assisted Learning, vol. 27, pp. 119–132, 2011.

[11] J. Keefe, "Learning style: an overview", NASSP's Student Learning Styles: Diagnosing and Prescribing Programs, pp. 1-17, 1979.

[12] T.K. Landauer and S.T. Dumais, "A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge", Psychological Review, vol. 104, pp. 211–240, 1997.

[13] C.D. Manning and H. Schütze, Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press, 1999.

[14] D.S. McNamara, A.C. Graesser, and M.M. Louwerse, "Sources of text difficulty: Across the ages and genres", in J. P. Sabatini, E. Albro and T. O'Reilly (Eds.): Measuring Up: Advances in How We Assess Reading Ability, pp. 89–116, Lanham, MD: R&L Education, 2012.

[15] G.A. Miller, "WordNet: A lexical database for English", Communications of the ACM, vol. 38, pp. 39–41, 1995.

[16] National Governors Association Center for Best Practices, Council of Chief State School Officers, "Common Core State Standards", Washington D.C., 2010.

[17] J. Nelson, C. Perfetti, D. Liben, and M. Liben, Measures of Text Difficulty: Testing Their Predictive Value for Grade Levels and Student Performance. Washington, DC: Council of Chief State School Officers, 2012.

[18] E. Page, "Analyzing student essays by computer", International Review of Education, 14(2), pp. 210–225, 1968.

[19] E. Popescu, "Providing collaborative learning support with social media in an integrated environment", World Wide Web, vol. 17(2), pp. 199-212, 2014.

[20] E. Popescu and F. Leon, "Analyzing the influence of learning styles on students' behavior in a social learning environment", Annals of the University of Craiova, Automation, Computers, Electronics and Mechatronics Series, vol. 11 (38) (1), pp. 29-38, 2014.

[21] C.E. Shannon, "Prediction and entropy of printed English", The Bell System Technical Journal, vol. 30, pp. 50–64, 1951.

[22] H. Slotnick, "Toward a theory of computer essay grading", Journal of Educational Measurement, vol. 9(4), pp. 253–263, 1972.

[23] B. Soloman and R.M. Felder, "Index of learning styles questionnaire", www.engr.ncsu.edu/learningstyles/ilsweb.html, 1998.

[24] E. Tang and C. Lam, "Building an effective online learning community (OLC) in blog-based teaching portfolios", Internet and Higher Education, vol. 20, pp. 79–85, 2014.

[25] S. Trausan-Matu, M. Dascalu, and P. Dessus, "Textual complexity and discourse structure in Computer-Supported Collaborative Learning", in Proc. 11th Int. Conf. on Intelligent Tutoring Systems (ITS 2012), pp. 352–357, Springer, 2012.

[26] W.T. Tsai, W. Li, J. Elston, and Y. Chen, "Collaborative learning using wiki web sites for computer science undergraduate education: A case study", IEEE Transactions on Education, vol. 54(1), pp. 114-124, 2011.

[27] T. A. van Dijk and W. Kintsch, Strategies of Discourse Comprehension. New York, NY: Academic Press, 1983.

[28] W. Wresch, "The imminence of grading essays by computer - 25 years later", Computers and Composition, vol. 10(2), pp. 45-58, 1993.

[29] Z. Wu and M. Palmer, "Verb semantics and lexical selection", in Proc. 32nd Annual Meeting of the Association for Computational Linguistics (ACL '94), pp. 133–138, 1994.