

Learning Analytics in MOOCs as an Instrument for Measuring Math Anxiety

Y. Y. Dyulicheva

Received in August 2021 **Yulia Y. Dyulicheva**, Candidate of Sciences in Mathematical Physics, Associate Professor, V.I. Vernadsky Crimean Federal University. Address: 4 Akademika Vernadskogo Ave, 295007 Simferopol. Email: dyulicheva_yu@mail.ru

Abstract In this paper, math anxiety descriptions are extracted from Massive Open Online Course (MOOC) reviews using text mining techniques. Learners' emotional states associated with math phobia represent substantial barriers to learning mathematics and acquiring basic mathematical knowledge required for future career success. MOOC platforms accumulate big sets of educational data, learners' feedback being of particular research interest. Thirty-eight math MOOCs on Udey and 1,898 learners' reviews are investigated in this study. VADER sentiment analysis, *k*-means clustering of content with negative sentiment, and sentence embedding based on the Bidirectional Encoder Representations from Transformers (BERT) language model allow identifying a few clusters containing descriptions of various negative emotions related to bad math experiences in the past, a cluster with descriptions of regrets about missed opportunities due to negative attitudes towards math in the past, and a cluster describing gradual overcoming of math anxiety while progressing through a math MOOC. The constructed knowledge graph makes it possible to visualize some regularities pertaining to different negative emotions experienced by math MOOC learners.

Keywords BERT, educational data mining, learning analytics, Massive Open Online Courses (MOOCs), math anxiety, text mining, VADER.

For citing Dyulicheva Y. Y. (2021) Uchebnaya analitika MOOC kak instrument analiza matematicheskoy trevozhnosti [Learning Analytics in MOOCs as an Instrument for Measuring Math Anxiety]. *Voprosy obrazovaniya/Educational Studies Moscow*, no 4, pp. 243–265 <https://doi.org/10.17323/1814-9545-2021-4-243-265>

With the rapidly growing number of mass online open courses (MOOC) and MOOC learners all over the world, huge sets of heterogeneous data have been accumulated, from age, status, and baseline level of knowledge to performance and progress through courses. Additional data may be obtained at the start and at the end of the course using questionnaires, unit tests, and final exams, and throughout the course by analyzing video viewing behavior, clickstreams, number of downloads, etc.

Learner–instructor interactions in MOOCs are being limited, learners can use reviews, forum comments, messengers, and social media to say what they think about course content and instructors. Analysis of learner reviews has given rise to a new field of learning analytics: educational data mining.

Mathematics MOOCs are often designed to improve basic mathematical knowledge. For instance, a project to develop a free open online course to help adults with low mathematical/statistical knowledge has been launched in the UK, outlining the universal requirements for online courses to explain the fundamental mathematical concepts in plain language [Griffiths et al. 2019]. Sometimes, math MOOCs are also created as scaffolding learning experiences to support students with low self-efficacy and high level of math anxiety, or as a tool to get students back into math after a holiday break [Lambert 2015]. MOOCs in mathematics can be applied to enhance teaching skills, serve as a network for exchanging effective teaching practices within the teachers' community [Taranto, Robutti, Arzarello 2020], and be adjusted to address the needs of specific target populations: individuals wishing to update their math skills; those who feel the need to acquire basic knowledge; and teachers who may use these resources with their students to develop teaching methodologies [Soares, Lopes 2016]. A distinct category of math MOOC learners is people with severe visual impairments and blindness [Kosova, Izetova 2020].

The recent years have seen a growing research interest in MOOC learning analytics, which is manifested in the emergence of MOOC datasets on data science platforms and Kaggle machine learning competitions (see kaggle.com). On Kaggle, for instance, one can find a dataset containing EEG brainwave data from college students while they watched MOOC video clips, a dataset with MOOC lecture data, etc.

1. Learning Analytics in MOOCs: Key Methodologies and Applications

The key methodologies in MOOC learning analytics include sentiment analysis, target group identification, analysis of content-based and clickstream data, course quality assessment, dropout prediction, and design of course and content recommender systems.

Sentiment analysis in MOOCs is often used to investigate students' attitudes, opinions, or emotions by detecting patterns in textual data and measuring their sentiment. Data obtained by analyzing learner feedback sentiment is used for identifying the reasons for attrition and lack engagement in MOOCs, developing strategies for improvement of MOOC content and teaching strategies, and getting a better understanding of student behaviors. Analysis of social media data (user profiles and comments) sheds light on MOOC learners' emotions and sentiment when progressing through a course. Learner feedback and forum posts are analyzed using hierarchical recurrent neural networks [Capuano et al. 2020]; two-polarity (positive/negative) sentiment analysis of MOOC reviews is performed using an ALBERT-BiLSTM mod-

el with three layers: word-embedding layer, semantic-extraction layer, and output layer [Wang, Huang, Zhou 2021]; collective sentiment from MOOC forum posts and its impact on student attrition are evaluated using survival analysis [Wen, Yang, Rose 2014]; social network analysis, cohort analysis, and identification of students who are actively participating in course discussions may assist in visualizing students' posting patterns in the course forum and building models of information diffusion [Sinha 2014]; and topic modeling is applied to trace discussion forum posts to MOOC content [Wong, Wong, Hindle 2019]. Another important task in educational data mining is to identify resource mentions in MOOC forum threads (sequence tagging), which is performed using the LSTM-CRF model [An et al. 2018].

Target group identification may help MOOC instructors develop strategies for better interaction with the course audience. Cluster analysis is often applied to solve learning analytics tasks of this type. Course satisfaction depends on the sentiment of target student groups which are identified using the VADER algorithm. In particular, it has been shown that most beginners are positive about MOOC content, while experienced participants often expect to learn about topics that are beyond the scope of the MOOC [Lundqvist, Liyanagunawardena, Starkey 2021]. Another criterion in target group identification is the differences in MOOC video learning behaviors, which serve the basis for content personalization [Zhang, Liu, Liu 2020].

Analysis of content-based and clickstream data allows detecting difficult or boring fragments of MOOC content and developing strategies for content personalization so as to improve the quality of study materials and resources. Learning analytics tasks in this subfield include using recurrent network modelling to predict the exact resource that a student will interact with next based on previous sequences of resource views and interactions in a MOOC [Tang, Peterson, Pardos 2016]; modelling student behaviors by analyzing video-watching clickstreams and sequences as well as other characteristics such as duration of video content viewing and in-video quiz completion rate [Brinton et al. 2015]. The data obtained can be used to build individual MOOC video interaction trajectories and develop strategies for personalized course recommendations.

Course quality assessment and development of reasonable assessment criteria constitute a challenging task, which can be solved, in particular, with the use of learning analytics tools. For instance, MOOC learning behavior and content perceptions can be studied by analyzing video traffic, forum posting, and the number of people who obtained the certificate [Luo et al. 2018].

Dropout prediction. Changes in student engagement throughout the course are monitored to develop retention strategies. One of the approaches to solving this task is to analyze learner activity and address the lack of feedback in a timely manner. Learner activity can be predicted, in particular, by obtaining weekly learning behavior statis-

tics using long short-term memory recurrent neural networks (LSTM-RNN) [Liu et al. 2018]. MOOC attrition can also be predicted by modeling representations of clicks and video [Jeon, Park 2020]. Furthermore, learning analytics tools are used for predicting student success as well [Bystrova et al. 2018].

Design of course and content recommender systems. As the number of available MOOCs is constantly growing, it becomes more and more difficult to select a course that would be affordable and matching the individual's needs. One of the approaches to developing course recommendations consists in constructing a binary tree of courses based on MOOC big data that allows identifying user preferences to find optimal solutions [Hou et al. 2016].

Development of big data mining algorithms in education has produced some tools for MOOC learning analytics. MOOCviz, a platform for visualizing data from edX and analyzing log data from Coursera, contains databases with information on students, including their activity and feedback. MOOCviz designers employed cohort and statistical analysis and used various heuristics to identify cohorts of students in MOOC courses based on resource use, country, etc. [Dernoncourt et al. 2013]. Other examples of MOOC analytical tools include PerspectivesX, a collaborative learning tool designed to support content and learner curation using topic modelling and deep learning techniques, and MessageLens, a visual analytics system to explore MOOC forum discussions [Bakharia 2017; Wong, Zhang 2018].

No studies of math anxiety in MOOC learners based on sentiment analysis of their reviews and comments can be found in the literature on MOOC learning analytics. This paper seeks to develop a methodology for detecting math anxiety based on analysis of math MOOC learners' feedback using machine learning techniques.

2. Approaches to Studying Anxiety

2.1. Math Anxiety

Mathematics anxiety is a serious problem associated with frustrating math learning experiences. The growing feelings of tension, failure, and disappointment can translate into resentment, fear, anxiety, chronic stress, and unwillingness to pursue careers requiring math knowledge and skills in the future. Math anxiety is understood as a specific emotional state of the student that triggers strong emotions such as hate and disgust and contributes to avoidance of any math-related experiences [Ashcraft, Moore 2009]. It is a widespread problem, affecting even engineering college students if they have difficulty learning the fundamental math concepts [Ma 1999]. Some scholars even qualify mathematics anxiety as a clinical pathology that impairs cognitive processes, contributes to social avoidance, and leads to negative emotional states even in well-performing students [Stella 2021].

There is empirical evidence of correlations between math anxiety and low mathematics achievement [Ashcraft, Moore 2009; Ma 1999]. In addition, mathematics anxiety may be exacerbated during the tran-

sition from elementary to secondary school or from secondary school to vocational or higher education due to changes in the learning environment [Field et al. 2021].

Survey response scales have been developed to measure math anxiety and discomfort experienced by math learners, e.g. the Abbreviated Math Anxiety Scale (AMAS) measuring anxiety in adolescents and adults and its modified versions mAMAS and EES-AMAS that measure math anxiety in children. The scales are based on evaluating emotional reactions toward math-related tasks [Carey et al. 2017; Priami et al. 2020].

Math anxiety prevention methods are built around the promotion of positive experiences, which is achieved by focusing the effort on adaptation with the use of innovative augmented and virtual reality teaching tools and a scaffolding system to support students in math problem solving [Dyulichева 2020].

Making allowance for the nature and sources of math anxiety is particularly important when developing mathematics MOOCs, as they imply no direct learner-instructor interactions. The prevalence of math anxiety is indirectly captured in the titles of some MOOCs. Udemy, for instance, offers a math course for beginners entitled *Calculus for Those Who Hate Calculus*, and an intermediate-level course under the title *Stress-Free Statistics for College and IB DP/AP Students: Mini-Course 2*.

Meanwhile, studies seeking to identify the right methods of MOOC design and teaching to prevent math anxiety as a result of online interactions have been extremely scarce. Diagnosis of math anxiety in MOOCs also remains an open question, as online learners are reluctant to participate in any additional questionnaires.

2.2. Text Mining for Anxiety Detection

Sentiment analysis of textual data (comments, reviews, social media profiles) has been successfully applied not only to measure the prevalence of negative sentiment in a community but also to diagnose depressive states, anxiety, and other mental disorders.

Depressive states are detected based on social media entries, using CollGram text analysis and sentiment analysis based on the Bidirectional Encoder Representations from Transformers (BERT) architecture. CollGram profiles include such measures as mutual information factor, t-score, and the number of idiosyncratic units describing painful reactions to some triggers in social media entries [Wołk, Chlasta, Hołas 2021]. Anxiety triggered by the COVID-19 pandemic was assessed through analysis of YouTube comments using various text vectorization techniques (Term Frequency—Inverse Document Frequency (TF-IDF), bag-of-words) and machine learning methods (Support Vector Machines (SVM), Random Forest, boosting, etc.) [Saifullah, Fauziah, Aribowo 2021]. Social media comments were analyzed using the Transformer-based Robustly Optimized BERT Pre-Training Approach (RoBERTa) model, LSTM neural networks, and BERT to classify five prominent kinds of mental illnesses: depression, anxiety, bipolar disorder, Atten-

tion Deficit Hyperactivity Disorder (ADHD), and Post Traumatic Stress Disorder (PTSD) [Murarka, Radhakrishnan, Ravichandran 2020]. Detection of depression and anxiety was performed on a set of 4,500 tweets using SVM with various text vectorization techniques and the BERT and ALBERT pre-trained language models [Owen, Camacho-Colados, Espinosa-Anke 2020].

Taking cue from the findings available [Wolk, Chlasta, Holas 2021; Murarka, Radhakrishnan, Ravichandran 2020; Stella 2021], we propose a methodology to analyze math anxiety by detecting math MOOC reviews with negative sentiment, clustering them, and visualizing the diagnosed states using a knowledge graph.

3. Dataset and Methods

Today, Coursera, Udemy, and EdX offer a variety of mathematics MOOCs with the most basic filtering options: by rating, price, video duration, skill levels, subtitles, etc. Figure 1 displays the number of English-taught math MOOCs on these three platforms found when searching for the keyword “math”. The largest number of math MOOCs for beginners and the smallest number of advanced-level courses is offered by Udemy.

For every skill level, we analyzed the MOOC titles and created word clouds representing the most frequent words (see Table 1). Generation of word clouds using the Python wordcloud library was preceded by preprocessing of the course titles, which involved removal of punctuation marks and stop words.

Two categories of English-taught beginner-level math courses on Udemy were analyzed: 27 courses with the keywords “fundamental”, “basic”, and “master” in the title, and 11 courses with the keywords “mental” and “vedic” in the title. Next, course reviews were scraped for both categories: 1,326 unique user reviews on the Fundamental/Basic Math Courses and 572 unique user reviews on the Mental/Vedic Math Courses.

As seen in Table 1, which is based on word frequency analysis, beginner-level math MOOCs on Udemy pay a lot of attention to the fundamentals of algebra, statistics, calculus, trigonometry, probability theory, graph theory, mental mathematics, and mathematics for machine learning. Coursera offers narrowly specialized math courses and courses with a focus on programming languages, machine learning (deep learning in particular), and data analysis. On EdX, beginner-level math courses are designed more to introduce learners into machine learning and quantum and classical mechanics, intermediate-level MOOCs teach fundamental mathematics (matrix algebra, linear algebra, differential calculus, etc.), mechanics, and electronics, and advanced-level courses focus on quantum calculus and applied problem solving.

Figure 2 displays key stages in the analysis of Udemy math course reviews and the detection of math anxiety based on negative feedback.

At the first stage, reviews are preprocessed by removing punctua-

Figure 1. The distribution of math MOOCs by skill levels, %.

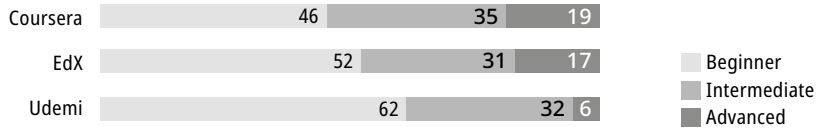
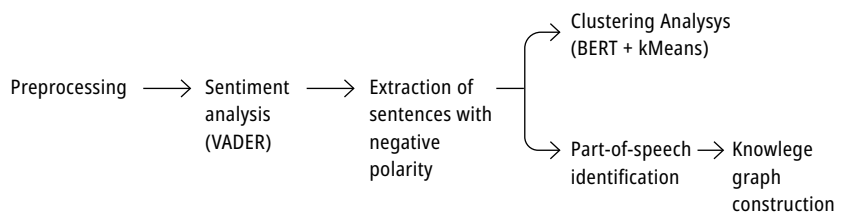


Table 1. Word clouds describing math MOOCs at different skill levels.

MOOC platform	Skill level		
	Beginner	Intermediate	Advanced
Udemy			
Coursera			
EdX			

Figure 2. Key stages in the analysis of math MOOC reviews.



tion marks and stop words, converting all words to lowercase, and tokenizing them using Python Natural Language Toolkit (NLTK).

At the second stage, sentiment analysis of reviews is performed with the help of the VADER algorithm. Sentiment is measured using the lexicon and rule-based approach or machine learning methods. VADER analysis based on rules and lexicons yields four polarities: positive, neutral, negative, and mixed. The advantage of applying the VADER tool is that it does not require using a dataset to teach the algorithm, and the disadvantage is that it overlooks words outside the sentiment lexicon when calculating the sentiment scores. The overall sentiment is described by the compound score, which is the normalized sum of valence scores calculated based on a particular heuristic and a sentiment lexicon. Normalization of the sum of valence scores from -1 (extremely negative sentiment) to $+1$ (extremely positive sentiment) is performed using the formula [Hutto, Gilbert 2014; Adarsh et al. 2019]:

$$\text{compoundScore} = \frac{x}{\sqrt{x^2 + \alpha}},$$

where α equals 15 by default, and x is the sum of all sentiment scores for the phrase (review).

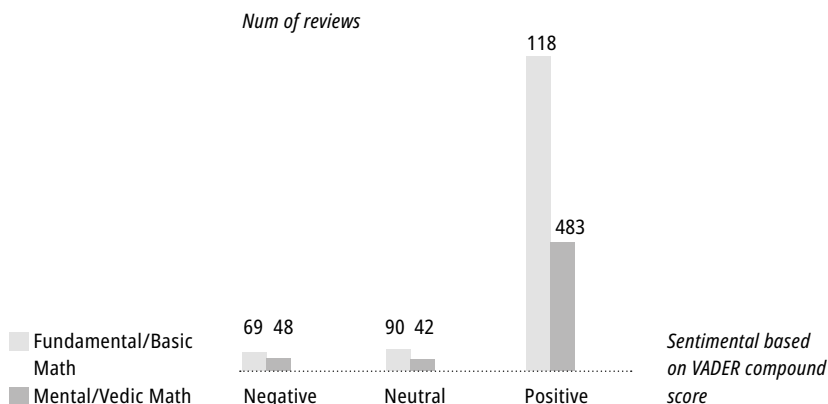
At the third stage, sentences with negative sentiment that contain no keywords but describe the course and/or instructor are identified using custom vocabularies, which comprise the words “course”, “lesson”, etc. and their synonyms, the words “instructor”, “tutor”, etc. and their synonyms, as well as instructors’ names (e. g. “Krista King”).

At the fourth stage, semantically related sentences describing math anxiety are identified and grouped using BERT, k -means clustering, and principal component analysis (PCA).

Representation of reviews as dense vectors of floating point values (embedding vectors) is performed using the pre-trained BERT model based on a bidirectional encoder neural network with the Transformer architecture [Devlin et al. 2019]. BERT demonstrates high accuracy and productivity on small datasets. After building vector representations of sentences with negative sentiment, the optimal number of clusters is determined using the elbow method and k -means clustering is performed according to the following procedure:

1. The optimal number of clusters k is used as the input data; k vector representations of math anxiety are randomly chosen as initial cluster means (centroids).
2. Each vector representation of a sentence describing math anxiety is assigned to the cluster with the least squared Euclidean distance, i. e. the one with the nearest mean.
3. According to the partitioning results, centroid coordinates are recalculated as the means of all vector representations assigned to each cluster.
4. Steps 2 and 3 are repeated until the assignments no longer change.

Figure 3. The initial distribution of sentiment (negative, neutral, and positive) for reviews about the Fundamental/Basic and Mental/Vedic Math MOOCs.



Combining BERT with *k*-means clustering allows identifying sets of sentences (clusters) based on structural specifics of the language.

Step 5 consists in identifying parts of speech and expression-like templates or patterns to capture relationships between different entities and construct a knowledge graph describing attitudes and emotional states extracted from the reviews.

4. Analyzing Student Feedback in Mathematics MOOCs to Detect Math Anxiety

4.1. Sentiment Analysis of Math MOOC Reviews

Below, we present the results of sentiment analysis of 1,326 reviews about the Fundamental/Basic Math Courses and 572 reviews about the Mental/Vedic Math Courses.

As it follows from the histogram in Figure 3, reviews about mathematics MOOCs are mostly positive. Sentiment analysis is performed using the VADER algorithm [Hutto, Gilbert 2014] and vaderSentiment library in Python.

The challenging part of math MOOC review sentiment analysis is that reviews with the overall positive or neutral sentiment sometimes contain one or more sentences describing frustrating experiences related to mathematics learning. A typical example of such review is presented in Table 2. The first row shows the original review about a math MOOC on Udemy, while the second row separates the fragment that contains no words describing the course or instructor. The columns display sentiment scores for the entire review (positive sentiment) and the separated fragment (negative sentiment) calculated in VADER.

Henceforth, we zero in on the negative feedback. By extracting sentences with negative sentiment that do not contain lexicon words describing the course or the tutor from the original dataset, we obtain

Table 2. An example of VADER sentiment analysis of a review about a math MOOC on Udemy.

Review/Fragment	Sentiment score			
	Positive	Neutral	Negative	Compound
I always despised Math throughout school because the teachers never made it fun. I almost didn't graduate HS because of my failure to attend Math class because I hated that much. After completing this course I definitely have a good foundation on fundamentals so this was not just a great refresher but I actually had to relearn all the concepts. Krista is an amazing teacher and I could only imagine how great she would be if she taught a live class. This course is packed with notes and practice tests to utilize in your learning and if you go through everything you will come out well equipped to learn the next level of Mathematics. I am moving on to Algebra next and Linear Algebra by Krista and who knows, if things go well I might go on to Calculus, Geometry and Probability & Stats. Krista has a really pleasant voice and she simplifies these concepts so well that even a child can grasp it. Highly recommended, there's a reason why her courses has the best reviews. Thanks Krista, you're awesome!	0.212	0.74	0.048	0.9913
I always despised Math throughout school because the teachers never made it fun.	0	0.677	0.323	-0.917

a dataset of 231 sentences with negative sentiment in the Fundamental/Basic category and 93 in the Mental/Vedic category.

4.2. Clustering of Negative MOOC Reviews

Of all the sentences with negative sentiment, those with the keywords “math”, “mathematic”, etc. are singled out on the basis of patterns, ending up with 52 sentences in the Fundamental/Basic category and 10 in the Mental/Vedic category. Clustering of these sentences is performed through vector representations using the BERT model and the *k*-means method. BERT produces dense vector representations of sentences describing math anxiety. The result is depicted in Figure 4.

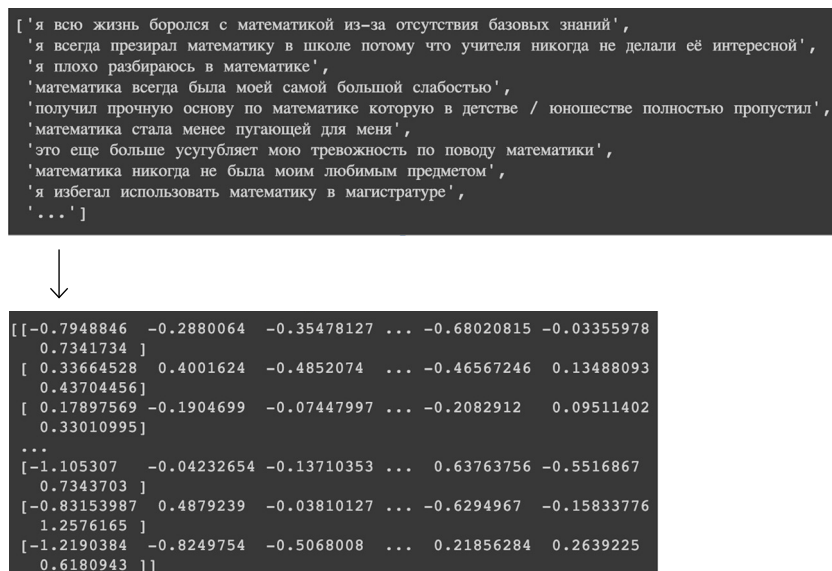
The combination of BERT and *k*-means clustering allows identifying clusters based on semantic similarity [Li et al. 2020]. The optimal number of clusters, required to apply the *k*-means algorithm, was determined using the elbow method and is equal to 5. Figure 5 shows the distribution of sentences describing math anxiety by clusters based on PCA and *k*-means clustering. Table 3 displays the clustering results.

Cluster 1 contains sentences that describe successful learning in the course despite prior frustrating math learning experience, using such words as “mental blocks” and “gaps in school knowledge”.

Cluster 2 includes descriptions of decreasing math anxiety, e.g. those that refer to math as “less intimidating” and “less confused”.

Cluster 3 consists of sentences expressing strong emotions about math-related problems, e.g. “math has always been my biggest weakness”, “math has always been my enemy”, “I always hated math with a passion”.

Figure 4. Sentence embedding in using BERT.



['i have struggled with math my whole life',
 'i always despised math throughout school because the teachers never made it fun',
 'i am not good at math',
 'math has always been my biggest weakness',
 'got solid foundation from math which in my childhood/teen missed totally',
 'math has become less intimidating now',
 'makes anxiety around math even worse',
 'math was never my favorite subject',
 'i avoided using any maths in my masters',
 '...']

Sentences in Cluster 4 describe how learners gradually overcome math learning problems as they are guided step by step into mathematics and acquire more and more skills within the course. For example, they include phrases like “grow in my math skills”, “the teacher teaching the math problems helps me understand how to work the problem”, “...it goes step by step and showing how the problem can be worked out”.

Cluster 5 features sentences that describe regrets about prior bad math learning experiences and missed opportunities, e. g. “It’s a pity that in my school days there wasn’t such a great teacher”, “...not being proficient in math caused me to fail a test”, “I almost didn’t graduate HS because of my failure to attend Math class because I hated that much”.

Figure 5. Distribution of sentences describing math anxiety by clusters based on PCA, *k*-means clustering, and BERT (Principal Component Analysis + *k*-means clustering)

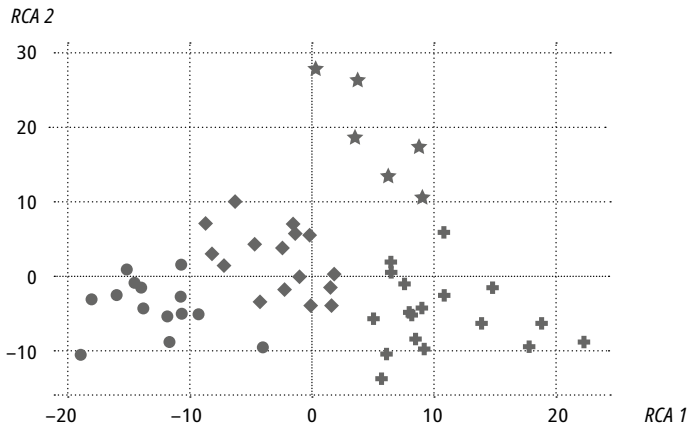


Table 3. Results of clustering of sentences with negative sentiment based on their semantic similarity.

Cluster No.	Cluster size	Random sentence example	Keywords	Mean compound sentiment score
1	7	I found I learnt quite a few new tricks that I wasn't taught in school	Mental blocks, gaps, school	-0.286
2	6	Math has become less intimidating now	Less, intimidating, confused, strange	-0.383
3	16	Math has always been my biggest weakness	Weakness, enemy, phobia, hate	-0.405
4	16	Each lesson challenged me and made me grow in my math skills	Challenge, skills, help, step by step, experience, negative, gradually	-0.326
5	17	It's a pity that in my school days there wasn't such a great teacher	Pity, wasn't, testing, school, negative	-0.498

Keywords in every cluster are determined using word frequency analysis.

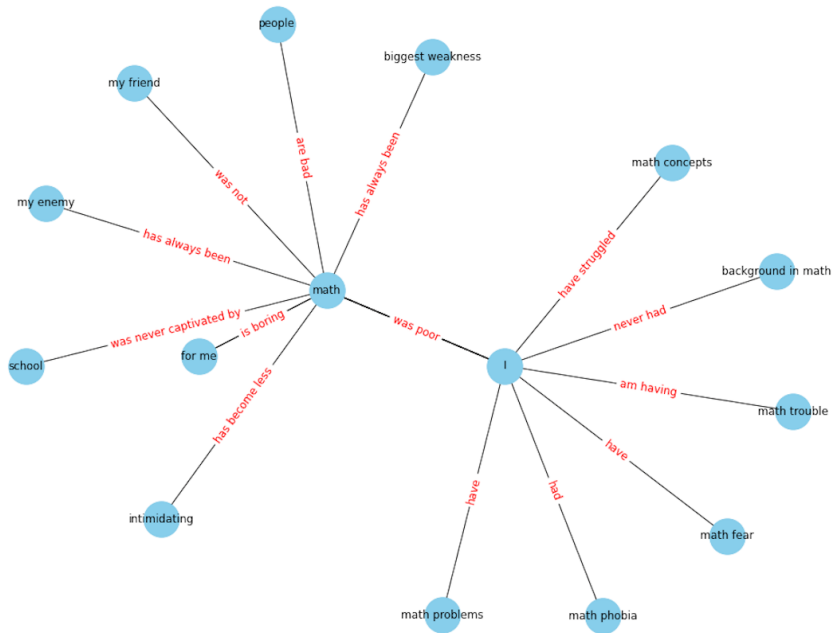
4.3. Constructing a Knowledge Graph Based on the Analysis of Negative Math MOOC Reviews

Knowledge graphs allow to visualize and structure relationships between entities as well as to describe their attributes. Nodes represent entities (documents, skills, job postings, tunes, etc.), and edges represent relationships (the Jaccard distance, events, etc.).

Knowledge graph and machine learning technologies are used in analysis of scientific publications [Chi et al. 2018]. Knowledge graphs can also be applied for mapping skills and matching them to job postings to facilitate labor market analysis [Groot de, Schutte, Graus 2021].

To explore math anxiety, we construct a knowledge graph based on part-of-speech identification and entity recognition using SpaCy

Figure 6. Knowledge graph constructed based on sentences with negative sentiment.



and NX Python libraries, respectively. For this purpose, we extract from the sentences with negative sentiment the pronoun “I” and the words “math”, “school”, etc. as well as the pattern consisting of the word “math” preceded or followed by a noun or pronoun that serve as subjects and objects defining the graph nodes. Relationships between the subject and the object are assigned to graph edges and labelled with verbs preceded or followed by an adverb or an adjective (if any).

The knowledge graph constructed based on the set of sentences with negative sentiment is shown in Figure 6. It allows visualizing the entities and relationships between them and facilitates result interpretation. In particular, it demonstrates negative emotions experienced by math learners—such as math phobia, apprehension about math, bad past experiences with math—and identifies their attitudes toward mathematics, e.g. “has always been” “is boring”, etc.

Visualization of entities describing math anxiety and relationships between them can be used by instructors, tutors, MOOC designers, and psychologists to analyze the sources of mathematics anxiety, search for ways to eliminate them, and render timely support to students.

5. Conclusions

Mathematics anxiety remains a serious problem hindering mathematical knowledge acquisition. This paper suggests a methodology for detecting math anxiety based on intelligent data analysis. In particular, the VADER algorithm for sentiment analysis was used to identify sentences with negative sentiment describing attitudes toward mathematics and math learning experiences at school; the BERT pre-trained language model was used to represent sentences with descriptions of math anxiety as vectors; the elbow method, k -means clustering, and principal component analysis were used to determine the optimal number of clusters, generate clusters of semantically related sentences describing math anxiety, and visualize them; and the part-of-speech identification and knowledge graph methods were used to visualize the relationships between learners and their emotions about past bad math learning experiences. The results can be used by MOOC instructors to improve the content of math courses, and by psychologists to develop recommendations on preventing and managing math anxiety disorders.

Despite the limited pool of math learner reviews with detailed descriptions of attitudes toward mathematics and specific courses and instructors, the findings obtained in this study can be used for diagnosing math anxiety and math phobia when working with large review datasets in various languages, and can be applied as an additional learning analytics tool in analyzing anxiety disorders in math learners.

This article was published with the support of the University Partnership Project run by National Research University Higher School of Economics.

References

- Adarsh R., Ashwin Patil, Shubham Rayar, Veena K. M. (2019) Comparison of VADER and LSTM for Sentiment Analysis. *International Journal of Recent Technology and Engineering*, vol. 7, iss. 6S, pp. 540–543.
- An Ya-H., Pan L., Kan M.-Ye., Dong Q., Fu Y. (2019) Resource Mention Extraction for MOOC Discussion Forums. *IEEE Access*, vol. 7, pp. 87887–87900, doi: 10.1109/ACCESS.2019.2924250
- Ashcraft M. H., Moore A. M. (2009) Mathematics Anxiety and the Affective Drop in Performance. *Journal of Psychoeducational Assessment*, vol. 27, no 3, pp. 197–205. doi:10.1177/0734282908330580
- Bakharia A. (2017) PerspectivesX: A Proposed Tool for Scaffold Collaborative Learning Activities within MOOCs. ArXiv preprint, abs/1704.04846
- Brinton Ch.G., Buccapatnam S., Chiang M., Poor H.V. (2015) Mining MOOC Click-streams: On the Relationship Between Learner Behavior and Performance. arXiv:1503.06489
- Bystrova T., Larionova V., Sinitsyn E., Tolmachev A. (2018) Uchebnaya analitika MOOK kak instrument prognozirovaniya uspehnosti obuchayushchikhsya [Learning Analytics in Massive Open Online Courses as a Tool for Predicting Learner Performance]. *Voprosy obrazovaniya/Educational Studies Moscow*, no 4, pp. 139–166. doi: 10.17323/1814-9545-2018-4-139-166
- Capuano N., Caballe S., Conesa J., Greco A. (2020) Attention-Based Hierarchical Recurrent Neural Networks for MOOC Forum Posts Analysis. *Journal of Ambient In-*

- telligence and Humanized Computing*, vol. 12, no 5, pp. 1–13. doi:10.1007/s12652-020-02747-9
- Carey E., Hill F., Devine A., Szucs D. (2017) The Modified Abbreviated Math Anxiety Scale: A Valid and Reliable Instrument for Use with Children. *Frontiers in Psychology*, vol. 8, Article no 11. doi: 10.3389/fpsyg.2017.00011
- Chi Ya., Qin Y., Song R., Xu H. (2018) Knowledge Graph in Smart Education: A Case Study of Entrepreneurship Scientific Publication Management. *Sustainability*, vol. 10, Article no 995. doi:10.3390/su10040995
- Dernoncourt F., Taylor C., O'Reilly U.-M., Veeramachaneni K., Wu Sh., Halawa Sh. (2013) MoocViz: A Large Scale, Open Access, Collaborative, Data Analytics Platform for MOOCs. Proceedings of the *NIPS Workshop on Data-Driven Education (Lake Tahoe, Nevada, USA, December, 9–10 2013)*. doi:10.13140/2.1.3749.1201
- Devlin J., Chang M.-W., Lee K., Toutanova K. (2019) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. arXiv: 1810.04805v2
- Dyulichева Yu. Yu. (2020) O primeneniya tekhnologii dopolnennoy i virtual'noy real'nosti v protsesse obucheniya matematike i fizike [About the Usage of the Augmented Reality Technology in Mathematics and Physics Learning]. *Open Education*, vol. 24, no 3, pp. 44–55. doi: 10.21686/1818-4243-2020-3-44-55
- Field A. P., Evans D., Bloniewski T., Kovas Yu. (2021) Predicting Maths Anxiety from Mathematical Achievement across the Transition from Primary to Secondary Education. *Royal Society Open Science*, vol. 6, Article no 191459. doi:10.1098/rsos.191459
- Griffiths L., Pratt D., Jennings D., Schmoller S. (2019) A MOOC for Adult Learners of Mathematics and Statistics: Tensions and Compromises in Design. *Topics and Trends in Current Statistics Education Research* (eds G. Burrill, D. Ben-Zvi), Cham, Switzerland: Springer International, pp. 351–371.
- Groot de M., Schutte J., Graus D. (2021) Job Posting-Enriched Knowledge Graph for Skills-Based Matching. arXiv:2109.02554v1
- Hou Yi., Zhou P., Wang T., Yu L., Hu Y., Wu D. (2016) *Context-Aware Online Learning for Course Recommendation of MOOC Big Data*. ArXiv, abs/1610.03147
- Hutto C.J., Gilbert E. (2014) VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the *Eight International AAAI Conference on Weblogs and Social Media (Ann Arbor, Michigan, USA, June 1–4 2014)*, pp. 216–225.
- Jeon B., Park N. (2020) Dropout Prediction over Weeks in MOOCs by Learning Representations of Clicks and Videos. ArXiv, abs/2002.01955
- Kosova Y., Izetova M. (2020) Dostupnost' massovykh otkrytykh onlayn-kursov po matematike dlya obuchayushchikhsya s ogranichennymi vozmozhnostyami zdorov'ya [Accessibility of Massive Open Online Courses on Mathematics for Students with Disabilities]. *Voprosy obrazovaniya/Educational Studies Moscow*, no 1, pp. 205–229. doi: 10.17323/1814-9545-2020-1-205-229
- Lambert S. (2015) Reluctant Mathematician: Skill-Based MOOC Scaffolds Wide Range of Learners. *Journal of Interactive Media in Education*, no 1. Article no 21. doi:10.5334/jime.bb
- Li B., Zhou H., He J., Mingxuan M., Yang Y., Li L. (2020) *On the Sentence Embeddings from Pre-trained Language Models*. arXiv:2011.05864v1
- Liu Z., Xiong F., Zou K., Wang H. (2018) *Predicting Learning Status in MOOCs Using LSTM*. arXiv:1808.01616v1
- Lundqvist K. O., Liyanagunawardena Th., Starkey L. (2021) Evaluation of Student Feedback within a MOOC Using Sentiment Analysis and Target Groups. *International Review of Research in Open and Distributed Learning*, vol. 21, no 3, pp. 140–156. doi:10.19173/irrodl.v21i3.4783
- Luo Yo., Li J., Xie Zh., Zhou G., Xiao X. (2018) MOOC Course Evaluation Based on Big Data Analysis. Advances in Computer Science Research. Proceedings of the *2018 International Conference on Computer Science, Electronics and Communication Engineering (CSECE2018) (Wuhan, China, February 7–8 2018)*, vol. 80, pp. 349–352. doi:10.2991/csece-18.2018.75

- Ma X. (1999) A Meta-Analysis of the Relationship between Anxiety toward Mathematics and Achievement in Mathematics. *Journal for Research in Mathematics Education*, vol. 30, no 5, pp. 520–540. doi:10.2307/749772
- Murarka A., Radhakrishnan B., Ravichandran S. (2020) Detection and Classification of Mental Illnesses on Social Media using RoBERTa. arXiv:2011.11226v1
- Owen D., Camacho-Collados J., Espinosa-Anke L. (2020) Towards Preemptive Detection of Depression and Anxiety in Twitter. Proceedings of the *Social Media Mining for Health Applications (Barcelona, Spain, Online, December 12 2020)*. arXiv:2011.05249
- Primi C., Donati M. A., Izzo V. A. et al. (2020) The Early Elementary School Abbreviated Math Anxiety Scale (the EES-AMAS): A New Adapted Version of the AMAS to Measure Math Anxiety in Young Children. *Frontiers in Psychology*, vol. 11, Article no 1014. doi:10.3389/fpsyg.2020.01014
- Saifullah S., Fauziah Yu., Aribowo A. S. (2021) *Comparison of Machine Learning for Sentiment Analysis in Detecting Anxiety based on Social Media Data*. Available at: <https://arxiv.org/ftp/arxiv/papers/2101/2101.06353.pdf> (accessed 20 October 2021).
- Sinha T. (2014) Supporting MOOC Instruction with Social Network Analysis. ArXiv, abs/1401.5175
- Soares F., Lopes A. P. (2016) Teaching Mathematics using Massive Open Online Courses. Proceedings of *10th International Technology, Education and Development Conference (Valencia, Spain, March, 7–9 2016)*, pp. 2635–2641.
- Stella M. (2021) Network Psychometrics and Cognitive Network Science Open New Ways for Detecting, Understanding and Tackling the Complexity of Math Anxiety: A Review. arXiv:2108.13800v1
- Tang S., Peterson J. C., Pardos Z. A. (2016) Modeling Student Behavior using Granular Large Scale Action Data a MOOC. ArXiv, abs/1604.04789
- Taranto E., Robutti O., Arzarello F. (2020) Learning within MOOCs for Mathematics Teacher Education. *ZDM: The International Journal on Mathematics Education*, vol. 52, no 2, pp. 1–15. doi:10.1007/s11858–020–01178–2
- Wang Ch., Huang S., Zhou Ya. (2021) Sentiment Analysis of MOOC Reviews via ALBERT-BiLSTM Model. *MATEC Web of Conferences*, 336, 05008. doi:10.1051/matec-conf/202133605008
- Wen M., Yang D., Rose C. P. (2014) Sentiment Analysis in MOOC Discussion Forums: What does it tell us? Proceedings of the *7th International Conference on Educational Data Mining, EDM 2014 (London, UK, July 4–7 2014)*, pp. 130–137.
- Wolk A., Chlasta K., Holas P. (2021) Hybrid Approach to Detecting Symptoms of Depression in Social Media Entries. Proceedings of the *Twenty-Fifth Pacific Asia Conference on Information System (Dubai, UAE, June 20–24 2021)*. Available at: <https://arxiv.org/ftp/arxiv/papers/2106/2106.10485.pdf> (accessed 20 October 2021).
- Wong A. W., Wong K., Hindle A. (2019) *Tracing Forum Posts to MOOC Content using Topic Analysis*. ArXiv, abs/1904.07307
- Wong J.-S., Zhang X. L. (2018) MessageLens: A Visual Analytics System to Support Multifaceted Exploration of MOOC Forum Discussions. *Visual Informatics*, vol. 2, iss. 1, pp. 37–49. doi:10.1016/j.visinf.2018.04.005
- Xhang F., Liu D., Liu C. (2020) MOOC Video Personalized Classification Based on Cluster Analysis and Process Mining. *Sustainability*, vol. 12, Article no 3066. doi:10.3390/su12073066