Sentiment Analysis Model Based on the Word Structural Representation

Gulmira Bekmanova[1,2] , Banu Yergesh[1(✉)] , and Altynbek Sharipbay[1,2]

[1] L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan
[2] Nuclear University MEPhI, Moscow, Russian Federation

Abstract. Over the past year, distance learning has become an integral part of our lives and a rapidly developing field. Compared to traditional classroom learning, online learning has advantages such as no place restrictions and a wide range of interactions. At the same time, distance learning lacks interaction between teachers and learners, and teachers cannot observe learners face to face. Identifying students' emotions during distance learning can have a positive impact on learning, improve learning outcomes and improve the effectiveness, quality of learning. This paper describes a theoretical model for determining sentiment/emotion from audio data based on speech recognition. The recognition model based on generalized transcription was proposed. This method can be used to determine the student's emotions during the distance learning and online exam.

Keywords: Natural language processing · Speech recognition · Emotion recognition · Sentiment analysis · Kazakh language · Distance learning

1 Introduction

Emotion recognition is the process of identifying human emotions. People differ from each other in their emotions. Human emotion recognition technology is a new area of research.

The latest advances in the artificial intelligence, deep learning methods, human-friendly Robotics, Cognitive Sciences are used to develop the field of affective computing and approach to creating emotional machines [1–3].

Currently, there are works on the automation of facial emotion recognition from video [4–8], by the rhythm of the voice from audio information [9–12], by the style of writing from texts [13–15].Based on various studies, emotions play a vital role in e-learning [5, 16, 17]. Likewise, improving the emotion recognition learning environment has been focused by researchers in the past few decades in the field of computer-based collaborative learning [18]. Teachers can change their teaching style according to the needs of the students.

Emotions act an important role in the analysis of a student's interest and learning outcomes in the course. Understanding emotions by facial expression is the fastest way to detect emotions [19, 20]. Results on sentiment analysis, emotion recognition in the Kazakh language published in [21–25].

Recently, distance learning has become a major formal in education. Distance learning has become a safe and viable option for lifelong learning due to COVID-19 pandemic.

In the new epidemic realities, the role of distance learning in education have greatly increased all over the world, a huge number of people have switched to remote work, schoolchildren and students study remotely. More than 91% of the world's students have been affected by school closures due to the COVID-19 pandemic, according to UNESCO. Even before the pandemic, the global e-education market was already seeing massive annual global growth. The mass transition of education to the distance format has become a serious challenge, both for the University, teachers, and for the students themselves.

Accordingly, the education system provisioning all students with same access to quality education during this crisis.

This gave a powerful impetus to the development of distance learning. According to the UNESCO study, most of the 61 countries surveyed have implemented some form of distance learning. The digital format of education is likely to become more popular in the post-pandemic period. Because this education format is effective and affordable [26, 27].

Our university switched to a distance learning format as well. During distance learning, the Microsoft Teams corporate platform is used, where online classes are recorded on video. And during the session, a proctoring system is used to pass computer testing. We have a database of video recordings of computer testing process. The video is recorded from the webcam and the audio from the microphone. During the examination, many students pronounce questions, talks to themselves, their speech have recorded, and their emotionality was kept by the proctors, so we decided to use the resulting audio material to analyze emotions. This paper describes a theoretical method for determining sentiment/emotion based on speech recognition. This method can be used to measure student emotions during distance learning and online examinations without geographic or cultural limitations.

2 Speech Recognition

Since this experimental work was carried out only to determine sentiment based on special words using generalized transcription recognition method. The recognition method based on generalized transcription, described in [28] used for general Kazakh word recognition before [29], for the first time it is proposed to apply this model for emotion recognition.

2.1 Structural Classification of Kazakh Words and Use of Generalized Transcriptions

This section represents some established statistical about Kazakh words structure. They are, as it seems to us, interesting by them and besides could serve a basis for using the generalized transcriptions. Let's divide all the symbols of the Kazakh alphabet into several natural classes:

W - аұыоеәүіөу
С - бвгғджзйлмнңр
F - сш
P - ққптфх

"W" – vowels plus the consonant "У", which, when pronounced, remains the vocal tract opened; "C" – voiced consonants; "F" – voiceless hush consonants; "P" – voiceless consonants, which when pronounced represent a pause in a word. Let us assume there is a significantly large dictionary of Kazakh words. Now it shall be a dictionary of initial forms containing 41791 words. Let's mark it out, replacing each symbol by the number of its class.

The words with the same marks shall be deemed to have a similar structure. Thus, the structure – is some model of gradation of vowels, consonants, hush sounds, etc. It appears that the number of Kazakh words with a similar structure is relatively small. For example, all the words with a structure WCCWFPWC are as follows:

алжасқан WCCWFPWC
алмастыр WCCWFPWC
ойластыр WCCWFPWC
үндескен WCCWFPWC
алдаспан WCCWFPWC

Here are the words with a structure WCWCWCPW:
ағарыңқы WCWCWCPW
амазонка WCWCWCPW
ұғыныңқы WCWCWCPW

Maximum number of words with the same structure CWCWC equal to 201, that is, And so on. Maximal number of words with a similar structure CWCWC is equal to 201, that amounts to about 0.5%. Moreover it is practically an exclusive case. All other structures contains much lesser words. We have proved it by the program which automatically marks out and selects words with the similar structure. Besides the selection of classes could be changed [30, 31].

The generalized transcription described is developed in the following way. First the voiceless consonants are distinguished from the recorded and processed sound signal as it was above described. The basis for it serves processing of a signal by a band pass filter with the range of transmission interval from 100 to 200 Hz as describes in the work, the coefficients of this filter are estimated according to the following equation:

$$a_k = a_{k,2} - a_{k,1} \tag{1}$$

where $a_{k,2}$ - is the coefficients of audio frequency band-pass filter which band-pass range is equal to 200 Hz, and $a_{k,1}$ - is the coefficients of audio frequency band-pass filter which band-pass range is equal to 100 Hz.

Audio frequency band-pass filters are estimated according to the following equation:

$$a_k = a_{-k} = \frac{2f_0}{f} \frac{\sin\left(k2\pi\frac{f_0}{f}\right)}{k2\pi\frac{f_0}{f}} \tag{2}$$

where k filter order, f_0 - is a transmission frequency of filter, f is a sampling frequency of signal.

The sounds noted differ from all the other ones by the fact that after such filtration their fragments become similar to a pause and contain a large number of constancy points. Thus, at these fragments the difference between the number of inconstancy points and the number of constancy points would be negative, that allows distinguishing them in the massive of such differences developed for the sequence of windows containing 256 counts.

Further the hush and pauselike sounds are selected in the resulted fragments. The analogue of total variation with variable upper limit is estimated:

$$V(0) = 0, \; V(n) = \sum_{i=0}^{n-1} |x_{i+1} - x_i| \tag{3}$$

Let N_1 - maximal number so that $V(N_1) \leq 255$.

N_2 - maximal number so that $W(N_1) \leq 255$ and so on. As a result the following massive of numbers appears

$$N_1, N_2 - N_1, N_3 - N_2 \ldots \tag{4}$$

At the hush segment the value (3) increases rapidly, i.e. the numbers (4) are relatively small. At the pause segment the value (3) increases slowly and consequently the numbers (4) are relatively large. To distinguish between the hush and the pause, introduce a threshold in the system, it is taken as 120.

When the hushes and the pauses are distinguished, the vowels and the voiced consonants are selected. The remaining fragments are divided into windows containing 256 counts and for them the value of total variation is calculated by Eq. (3), then the average of these values is estimated which is considered as the limit. All the values that are above the average are marked "B", below the average – "H". Then the interval, which the described procedure is conducted at, is moved one window to the right and the procedure is repeated. It continues up to the moment when the end of the interval falls outside the boundaries of the fragment (Fig. 1).

The marks of segmentations are entered at the points where the symbols "H" are changed into "B", or "B" into "H". B-fragment is deemed corresponding to the vowel (the symbol W is put near the left mark). H-fragment is deemed corresponding to the voiced consonant (the symbol C is put near the left mark) [30–32].

This algorithm could help to develop even more generalized transcription, namely to divide all the sounds into 2 natural classes: vowels and consonants. Such division gives good results in little dictionaries as well.
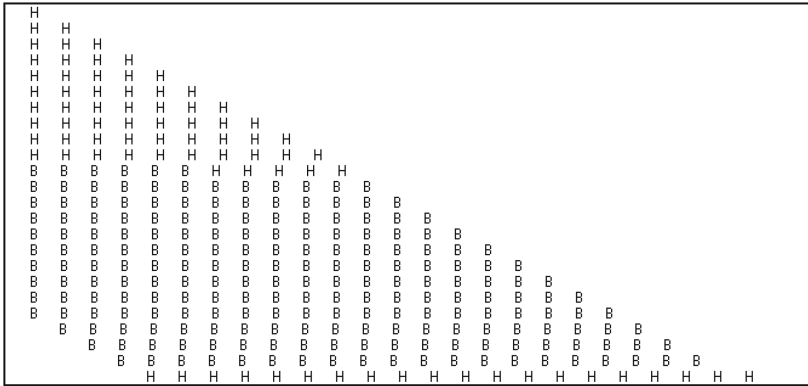
```
H
H   H
H   H   H
H   H   H   H
H   H   H   H   H
H   H   H   H   H   H
H   H   H   H   H   H   H
H   H   H   H   H   H   H   H
H   H   H   H   H   H   H   H   H
H   H   H   H   H   H   H   H   H   H
B   B   B   B   B   B   H   H   H   H   H
B   B   B   B   B   B   B   B   B   B   B   B
B   B   B   B   B   B   B   B   B   B   B   B   B
B   B   B   B   B   B   B   B   B   B   B   B   B   B
B   B   B   B   B   B   B   B   B   B   B   B   B   B   B
B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B
B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B
B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B
B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B
B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B
    B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B
        B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B
            B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B   B
            H   H   H   H   H   H   H   H   H   H   H   H   H   H   H   H   H   H   H   H   H   H
```

Fig. 1. Signal segmentation

2.2   Construction of Generalized Transcriptions of Sentiment Dictionary

The dictionary of generalized transcriptions of the sentiment dictionary obtained in [34] allows us to construct its generalized transcriptions and apply them to search for emotionally colored words.

Example of emotional words in the Kazakh language with generalized transcription:

Table 1. Example of sentimental words with generalized transcription

Sentiment words	Generalized transcriptions
onai (easy)	WCWC
quanyshtymyn (I'm glad)	PWWCWFPWCWC
senimdimin (I'm sure)	FWCWCCWCWC
daiyndaldym (prepared)	CWCWCCWCCWC
bilemin (I know)	CWCWCWC
aqymaq (stupid)	WPWCWP
qobalzhimyn (I'm worried)	PWCWCCWCCWC
bilmeimin (I don't know)	CWCCWCCWC
qorqamyn (I'm afraid)	PWCPWCWC
qiyn (difficult)	PWCWC

3   Sentiment Analysis

3.1   Dataset

Table 1 below shows the number of videos with audio recordings of the exam process (Table 2).

Table 2. Number of videos with audio recordings

Academic period	Number of videos with audio recordings
1 Semester	14606
2 Semester	4732
Total	19338

3.2   Sentiment Detection Process

The recorded speech is recognized, and then the text is analyzed to determine the sentiment. The process described in Fig. 2. Models and methods for determining the sentiment of texts in the Kazakh language are described in the works [22, 24, 33, 35].
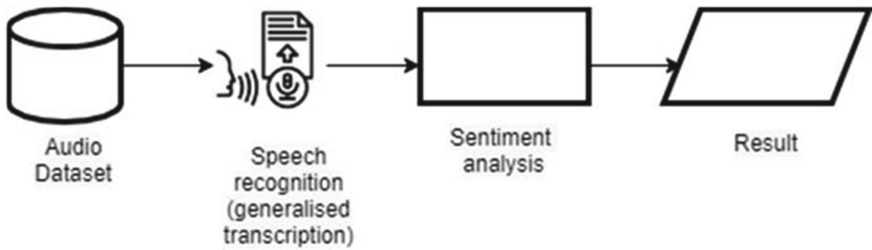

Fig. 2. Sentiment detection process

After speech recognition, words that express emotions are highlighted and tagged into 3 classes Positive, Neutral and Negative (Table 3). This tagged dataset will be used as test set.

Table 3. Sentiment classes

Polarity	Emotion	Number of audio recordings
Positive	Happy	1502
Neutral	Calm, neutral	1000
Negative	Sad, angry, fearful	500

Example:

Recognized words	Sentiment class
senimdimin (I'm sure)	Positive
daiyndaldym (prepared)	Positive
bilemin (I know)	Positive
aqymaq (stupid)	Negative
qobalzhimyn (I'm worried)	Negative

4   Conclusion

This work is devoted to the study and solution of the problem of sentiment analysis of students' speech. As a result of the study of models and methods of speech sentiment analysis of Kazakh language, a semantic base of emotional words with generalized transcription was obtained. Emotion detection models of audio sounds recorded during the distance exam were proposed and implemented. Experimental works in progress now. The implementation and application of this model can improve the interaction between the teacher and the student, improve the quality of distance learning, and help to personalize education. We are planning to complete experimental work and compare with other state-of-the-art methods. In the future, it is planned to study video files and determine the emotion from the image and video.

References

1. Franzoni, V., Milani, A., Nardi, D., Vallverdú, J.: Emotional machines: the next revolution. Web Intell. **17**(1), 1–7 (2019). https://doi.org/10.3233/WEB-190395
2. Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A., Cambria, E.: DialogueRNN: an attentive RNN for emotion detection in conversations. In: Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, pp. 6818–6825 (2019). https://doi.org/10.1609/aaai.v33i01.33016818
3. Biondi, G., Franzoni, V., Poggioni, V.: A deep learning semantic approach to emotion recognition using the IBM Watson Bluemix Alchemy language. In: Gervasi, O., et al. (eds.) ICCSA 2017. LNCS, vol. 10406, pp. 718–729. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62398-6_51
4. Stappen, L., Baird, A., Cambria, E., Schuller, B.W., Cambria, E.: Sentiment analysis and topic recognition in video transcriptions. IEEE Intell. Syst. **36**(2), 88–95 (2021). Article no. 9434455
5. Yang, D., Alsadoon, A., Prasad, P.W.C., Singh, A.K., Elchouemi, A.: An emotion recognition model based on facial recognition in virtual learning environment. Procedia Comput. Sci. **125**, 2–10 (2018)
6. Gupta, O., Raviv, D., Raskar, R.: Deep video gesture recognition using illumination invariants. ArXiv abs/1603.06531 (2016)

7. Kahou, S.E., et al.: Combining modality specific deep neural networks for emotion recognition in video. In: Proceedings of the 15th ACM on International Conference on Multimodal Interaction, pp. 543–550. ACM (2013)

8. Ozdemir, M., Elagoz, B., Alaybeyoglu, A., Akan, A.: Deep learning based facial emotion recognition system. In: 2020 Medical Technologies Congress (TIPTEKNO), pp. 1–4 (2020). https://doi.org/10.1109/TIPTEKNO50054.2020.9299256

9. Franzoni, V., Biondi, G., Milani, A.: Emotional sounds of crowds: spectrogram-based analysis using deep learning. Multimedia Tools Appl. **79**(47–48), 36063–36075 (2020). https://doi.org/10.1007/s11042-020-09428-x

10. Salekin, A., et al.: Distant emotion recognition. Proc. ACM Interact. Mob. Wear. Ubiquit. Technol. **1**(3), 1–25 (2017). https://doi.org/10.1145/3130961

11. Fayek, H.M, Lech, M, Cavedon, L.: Towards real-time speech emotion recognition using deep neural networks. In: Proceedings of the 9th International Conference on Signal Processing and Communication Systems, ICSPCS 2015, pp. 1–5 (2015). https://doi.org/10.1109/ICSPCS.2015.7391796

12. Mirsamadi, S., Barsoum, E., Zhang, C.: Automatic speech emotion recognition using recurrent neural networks with local attention. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing (2017). https://doi.org/10.1109/ICASSP.2017.7952552

13. Franzoni, V., Biondi, G., Milani, A.: A web-based system for emotion vector extraction. In: Gervasi, O., et al. (eds.) ICCSA 2017. LNCS, vol. 10406, pp. 653–668. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62398-6_46

14. Franzoni, V., Li, Y., Mengoni, P.: A path-based model for emotion abstraction on Facebook using sentiment analysis and taxonomy knowledge. In: Proceedings of International Conference on Web Intelligence, WI 2017, Leipzig, pp. 947–952 (2017)

15. Canales, L, Martinez-Barco, P.: Emotion detection from text: a survey. In: Processing of the 5th Information Systems Research Working Days (JISIC 2014), pp. 37–43 (2014)

16. Immordino-Yang, M.H., Damasio, A.: We feel, therefore we learn: the relevance of affective and social neuroscience to education. Mind Brain Educ. **1**(1), 3 (2007). https://doi.org/10.1111/j.1751-228X.2007.00004.x

17. Durães, D., Toala, R., Novais, P.: Emotion analysis in distance learning. In: Auer, M.E., Rüütmann, T. (eds.) ICL 2020. AISC, vol. 1328, pp. 629–639. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68198-2_58

18. Baker, M., Andriessen, J., Järvelä, S.: Affective Learning Together. Social and Emotional dimension of collaborative learning. Routledge, Abingdon (2013)

19. Krithika, Lb., Lakshmi, G.G.: Student emotion recognition system (SERS) for e-learning Improvement based on learner concentration metric. Procedia Comput. Sci. **85**, 767–776 (2016). https://doi.org/10.1016/j.procs.2016.05.264

20. Franzoni, V., Biondi, G., Perri, D., Gervasi, O.: Enhancing mouth-based emotion recognition using transfer learning. Sensors 5222 (2020). https://doi.org/10.3390/s20185222

21. Yergesh, B., Bekmanova, G., Sharipbay, A., Yergesh, M.: Ontology-based sentiment analysis of Kazakh sentences. In: Gervasi, O., et al. (eds.) ICCSA 2017. LNCS, vol. 10406, pp. 669–677. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62398-6_47

22. Yergesh, B., Bekmanova, G., Sharipbay, A.: Sentiment analysis of Kazakh text and their polarity. Web Intell. **17**(1), 9–15 (2019). IOS Press. https://doi.org/10.3233/WEB-190396

23. Zhetkenbay, L., Bekmanova, G., Yergesh, B., Sharipbay, A.: Method of sentiment preservation in the Kazakh-Turkish machine translation. In: Gervasi, O., et al. (eds.) ICCSA 2020. LNCS, vol. 12250, pp. 538–549. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58802-1_38

24. Yergesh, B., Bekmanova, G., Sharipbay, A.:Sentiment analysis on the hotel reviews in the Kazakh language. In: Proceedings of 2nd International Conference on Computer Science and Engineering (UBMK), Antalya, pp. 790–794 (2017)

25. Bekmanova, G., Yelibayeva, G., Aubakirova, S., Dyussupova, N., Sharipbay, A., Nyazova, R.: Methods for analyzing polarity of the Kazakh texts related to the terrorist threats. In: Misra, S., et al. (eds.) ICCSA 2019. LNCS, vol. 11619, pp. 717–730. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24289-3_53

26. Facts and Stats that Reveal the Power of eLearning. https://www.shiftelearning.com/blog/bid/301248/15-facts-and-stats-that-reveal-the-power-of-elearning. Accessed 01 May 2021

27. Online Education Statistics: 2020 Data on Higher Learning & Corporate Training. http://www.guide2research.com/research/online-education-statistics. Accessed 01 May 2021

28. Shelepov, V.Yu., Nitsenko, A.V.: On the recognition of Russian words using generalized transcription. Probl. Artif. Intell. **1**(8), 50–56 (2018). (in Russian)

29. Sharipbayev, A.A., Bekmanova, G.T., Shelepov, V.Yu.: Formalization of phonologic rules of the Kazakh language for system automatic speech recognition. https://kze.docdat.com/docs/411/index-1914530.html. Accessed 01 June 2021

30. Nitsenko, A.V., Shelepov, V.: Algorithms for phonemic recognition of words for a given dictionary. Artif. Intell. [Iskusstvennyy intellekt] **4**, 633–639 (2004). (in Russian)

31. Shelepov, V.Yu.: The concept of phonemic recognition of separately pronounced Russian words. Recognition of syntactically related phrases. In: Materials of International Scientific-Technical Conference "Artificial Intelligence", Donetsk-Taganrog-Minsk, pp. 162–170 (2007). (in Russian)

32. Shelepov, V., Nitsenko, A.V.: To the problem of phonemic recognition. Artif. Intell. [Iskusstvennyy intellekt] **4**, 662–668 (2005). (in Russian)

33. Yergesh, B., Sharipbay, A., Bekmanova, G., Lipnitskii, S.: Sentiment analysis of Kazakh phrases based on morphological rules. J. Kyrgyz State Tech. Univ. named after I. Razzakov **2**(38), 39–42 (2016). Bishkek

34. Yergesh, B.Zh.: Sentiment determination of the Kazakh language texts based on the dictionary of emotional vocabulary. In: Proceedings of 5th International Conference on Computer Processing of Turkic Languages "TurkLang 2017", vol. 1, pp. 62–67. Publishing House of the Academy of Sciences of the Republic of Tatarstan, Kazan (2017). (in Russian)

35. Yergesh, B., Sharipbay, A., Bekmanova, G.: Models and methods of sentiment analysis of texts in the Kazakh language. In: Computational Processing of the Kazakh Language: Collection Of Scientific Papers, Chapter 5. Kazakh University, Almaty (2020). (in Russian)