Exploring Text-based Emotions Recognition Machine Learning Techniques on Social Media Conversation

Sherzod Yarashev[1*], Bekzod Erkinov[1], Elyor G'aybulloyev[1], Jurabek Abdiyev[2], Nigora Abdiyeva[3], Ruslan Malikov[3]

[1]Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan, Tashkent, Amir Temur street 108.
[2]Physical-technical Institute of NPO "Physics – Sun" of Uzbekistan Academy of Sciences Uzbekistan, Tashkent, Chingiz Aitmatov street 2B.
[3]Samarkand State Institute of Foreign Languages. Samarkand, 140104, street Bustonsaroy 93
Author: sherzodyarashev1997@gmail.com (Sh. Yarashev)

Abstract: *Emotions hold a paramount role in the conversation, as it expresses context to the conversation. Text/word in conversation consists of lexical and contextual meanings. Extracting emotions from text has been an interesting work recent these years. With the advancement of machine learning techniques and hardware to support the machine learning process, recognizing emotions from a text with machine learning provides promising and significant results. This research aims to explore several popular machine learning to recognize emotions from a conversation in social media. The algorithms proposed in this research are ranged from traditional machine learning to deep learning techniques. The dataset used in this paper is provided by Affective Tweets, with a baseline of F1S core of 0.71 with word N-grams and Senti Strength. The research contributes extensive explorations in a number of machine learning algorithms, resulting in a total of 2302 features sets were explored, where each features sets has 100-1000 features extracted from the text. The results demonstrate Generalized Linear Model provides the best Accuracy score (0.92), Recall (0.902), Precision (0.902), F1 score (0.901) with standard deviation of accuracy of ±1, 2%.*

Keywords*:* Machine Learning; Emotions Recognition, Social Media, Text-based Emotions

1. INTRODUCTION.

Emotions play an essential role in the conversation. They express meanings to the conversation reciprocally with the context of the text. Recognizing emotions form text has been an attractive task recent these years. With the advancement of machine learning techniques and hardware to support the machine learning process, recognizing emotions from a text with machine learning provides promising and significant results. The task of recognizing emotions seems a straightforward task for a human. However, it seems a cumbersome task for a social-ignorant computer[1]. Several aspects need to be comprehended to build a social computer that not only capable of understanding the human verbal meaning, but also competent to perceive the conversation intent from the non-verbal cues (e.g. emotions from facial expressions, voice prosody, and meaning from the particular word). The ability to perceive non-verbal context from the conversation will make the interaction more colourful[2,1]. Machine learning can be implemented to teach the machine on how to understand the context of emotions from the conversation. In order to deliver outstanding results on the training, the data provided as the fuel in the machine learning training should contain a colorful conversation. One of the potential resources for dataset is social media. Social media presents a colorful and natural conversation, where people can express their argument, emotions, opinions, and stories[3].

This research aims to explore a number of machine learning algorithms to recognize emotions from the conversation in social media. The dataset used in this research is from the AffectiveTweets[4], where it provides a total of more than 7000 emotions labelled utterances. The algorithms proposed in this research are ranged from the traditional machine learning to deep learning techniques, they are Naive Bayes, Generalized Linear Model, Fast-Large Margin, Artificial Neural Network, Decision Tree, Random Forest, and Support Vector Machine. The baseline used in this paper is from the one provided by Affective Tweets (*F*1*S core* = 0.71), with Word N-grams, All lexicons + Senti Strength as the features[4]. The research contributes to the exploration of several machine learning algorithms with a large number of analysis in the combination of settings in each algorithm. A total of 2302 features sets were explored, where each features sets has 100-1000 features extracted from the text. The results demonstrate Generalized Linear Model provides the best Accuracy score (0.92), Recall (0.902), Precision (0.902), F1 score (0.901) with standard deviation of accuracy of ±1, 2%. Moreover, Decision Tree and Random Forest were not suitable for this problem in this research. The rest of the paper is organized as follows: the literature and recent work related to this research are described in the next section. The proposed methods and the settings of each algorithm are thoroughly in the Methods section. The results are discussed in the Results and Discussion section. Finally, the implication of the research and the future direction of the research is illustrated in the final section.

2. RECENT WORK.

In this section, we discuss accessible machine learning and deep learning methods mentioned in Chapter 1. In the past few decades, there are various algorithms which can identify emotions from natural language text. It can be divided into three major approaches: pattern-matching, machine learning, and deep learning. The work of[5] and[6] captures users' intention located in the natural-language text query format by using pattern-matching approach. The use of pattern-matching method is easy-to-use and straightforward. Nevertheless, words can have different meanings; causing the incorrect result of emotion annotations. In their paper, Yasmina et al.[7] identify emotions from YouTube comments using unsupervised machine learning algorithm, Support Vector Machine. It utilizes 200,000 comments from various video categories extracted by using YouTube API. The algorithm works on word-level, which then combined into emotion classification at the sentence level. It results in 92.75% as average precision and 68.82% as average accuracy. Furthermore, Banik and Rahman[8] proposed emotions models from movie reviews with Naive Bayes and Support Vector Machine with N-Gram as the features. As the pre-processing phase, stemming was also applied to the text to extract the base word. The best results achieved in this research were the F1 score of 86.00% and 83.00% for Support Vector Machine and Naive Bayes, respectively. Hasan et al.[9] proposed Naıve Bayes, Support Vector Ma-chine, and Decision Tree to recognize emotions from 135,000 processed labelled text from Twitter. The features used in this research were Unigram of word, emoticons, punctuation, and negation. The best results of F1 score achieved in this research were 90.00% for Support Vector Machine (only unigram feature), 90.00% for Decision Tree (using all features), and 90.00% for Naive Bayes (with all features). Liu et al.[10] also proposed Extreme Learning Decision Tree, Support Vector Machine, and Back-propagation Neural Network for emotions recognition. The emotions were classified into six basic emotions. The results are 89.60%, 87.20%, 82.30% for Extreme Learning Decision Tree, Support Vector Machine, and Back-propagation Neural Network, respectively, as the average classification accuracy. Zhang et al.[11] also proposed multi-label learning model to classify emotions from an online social network with results of 51.90, 57.90, 51.40, and 57.70 for classification accuracy, precision, recall, and F1-Score respectively. Moreover, currently, the application of advanced neural network such as deep learning algorithm used in this problem[12,13]. Zheng and Yang[13] recently explored several traditional algorithms, such as Naıve Bayes, Support Vector Machine, K-Nearest Neighbors as well as deep learning algorithms such as Back-propagation Neural Network, Convolutional Neural Net-work, and Long-Short Term Memory architectures. The research classified two classes of emotions (i.e. positive and negative) with the positive class performed better than the negative class. The best classification accuracy was archived by Long-Short Term Memory with the accuracy of 76.10% and 51.20% for positive and negative class, respectively. In comparison, the lowest performance belonged to the K-Nearest Neighbors algorithm with a classification accuracy of 67.70% 45.70% for positive and negative class, respectively.

In their research, Wang et al.[14] build a predictive analytic model to identify restaurants which need health inspection. It uses 235,000 online reviews from 5,800 from Yelp users as corpus. Nevertheless, the dataset is imbalanced and illness-related features are hardly found. For the language features, N-gram[15] and sentiment with lexicons are used based on LIWC[16] dictionary. In the prediction models, five algorithms are compared: generalized linear models (*GL*), naive bayes (*NB*), support vector machines (*SVM*), random forest (*RF*), and recurrent neural network (*ANN*). Based on the experiment, all prediction models outperform the *GL* method. The best accuracy goes to *NB* and the best precision goes to *SVM*. Furthermore, all recall values are less than 70% because of imbalance data.

3. METHODS.

Seven most accustomed machine learning algorithms were explored to train the AffectiveTweets dataset[4]. Table 3 demonstrates the dataset profile. The dataset consists of a total of 7102 labelled English utterances crawled from social media. The column "Class" represents the class of the emotions from the text (i.e. Anger, Sadness, Fear, and Joy). Each class has a quite imbalanced number of the utterances, with the largest number belong to Fear (2252, 32%), and the lowest number belong to the Sadness (1701, 22%, see "Actual" and "% each class" Column). To make the class balance, we performed up-scaling, and down-scaling sampling to the data resulted in a total of 6801 texts where each class has balanced number of text (1700 each, see "Sampling" column). The dataset then divided into 85%:15% of training and testing set (see "Train" and "Test" Column). The "% changes" column indicates the percentages of labelled data changes after sampling compared to before sampling. Before throwing the dataset to the various algorithms, the data was pre-processed to increase the quality of the data. The pre-processing steps applied were: removing noisy data, tokenization, filtering stop-words, filtering tokens based on length, stemming, transforming cases, and extracting text features from the pre-processed data.

Table 1. The Dataset Profile

Class	Actual	% each class	Sampling	% changes	Train	Test
Anger	1701	24%	1701	0%	1446	255
Sadness	1533	22%	1700	11%	1445	255
Fear	2252	32%	1700	-25%	1445	255
Joy	1616	23%	1700	5%	1445	255
TOTAL	7102	100%	6801	-4%	5781	1020

Not meaningful and irrelevant text, such as header, HTML, XML (e.g. JSON) tags were removed in the first step. Next, the sentences were converted into a token, a small block of words. Commonly used words (e.g. I, am, in, the) were also filtered and removed. To increase the training process, tokens with a length of more than 15 characters were removed. Next, the root or base of the words was extracted in the stemming process. Finally, as an initial setting, 1000 vector of features were extracted from the text to represent the whole dataset. To convert the bag of words from the dataset into a vector space, we implemented Term Frequency, Term Occurrences and Term Frequency inverted document frequency (TF-IDF) vectorisation[17]. Seven algorithms proposed in this research are: Naive Bayes, Generalized Linear Model, Fast-Large Margin, Artificial Neural Network, Decision Tree, Random Forest, and Support Vector Machine. For each algorithm proposed in this research, we explore a total of 2302 feature sets to find the optimal trade-offs between complexity and error rates. The best feature sets should have low complexity and error rates. Moreover, not all 1000 features were used depends on the algorithm. The goal is to minimize the complexity and achieve the best result at the same time. Table 3 demonstrates the overview of feature sets exploration. The process was done automatically and will stop when either all combinations were explored or a certain time has passed (i.e. 24 hours for each algorithm). The largest feature sets explored in Support Vector Machine (1449 features sets) with only 250 features used.

Table 2. Automatic Features Sets Exploration

Algorithm	Evaluated Feature Sets	Features Used
Naıve Bayes	102	700
Generalised Linear Model	102	650
Fast Large Margin	102	650
Artificial Neural Network	116	1000
Decision Tree	116	550
Random Forest	315	600
Support Vector Machine	1449	250

Next step of this research was to train the pre-processed data to all the algorithms proposed. In the training process, we also explored several combinations of the hyper-parameters to optimize the results in each algorithm except for Artificial Neural Network. The detailed of the results and discussions of the hyper-parameters optimization are discussed in the next section (see Table 4 - 7). Table 3 illustrates the hyper-parameters settings for Artificial Neural Network algorithm used in this research. The architecture consists of four layers with the input layer has 1000 Rec-tidier units/nodes, represents the number of features used. Second and third layers act as the hidden layers with fifty Rectifier units/nodes in each layer. The fourth layer represents the output of classification with 4 Softmax units/nodes that represent Anger, Sadness, Fear, and Joy emotions. There is no dropout performed in all the layers. L1 and L2 Normalization value for all layers are 0, 000010, and 0 respectively. As the final process, the best-trained model in each algorithm was evaluated with a total of 1020 text with 255 texts in each class. The results are comprehensively discussed in the next section.

Table 3. Artificial Neural Network Hyper parameter

Layer	Unit	type	Dropout	L1	L2
1	1000	Rectifier	0%	0.000010	0
2	50	Rectifier	0%	0.000010	0
3	50	Rectifier	0%	0.000010	0
4	4	Softmax	0%	0.000010	0

4. RESULTS AND DISCUSSIONS.

In the training process, a total of 5781 labelled text (85% of the dataset) were used in several machine learning algorithms. The initial features extracted from the dataset is 1000; however, the number of features used in each algorithm depends on the exploration of the features sets combination (see Table 3. The features sets explorations aimed to find the optimal trade-offs between complexity and error rates (i.e. low complexity and low error rates). The best features set was then used as the fuel of emotions recognition modelling using machine learning. The algorithms used in this research are: Naive Bayes, Generalized Linear Model, Fast-Large Margin, Artificial Neural Network, Decision Tree, Random Forest, and Support Vector Machine. The best model in each algorithm was tested with a total of 1020 text in the dataset. The results are shown in table 8. In the training process, some hyper-parameters combinations also explored in Fast-Large Margin, Decision Tree, Random Forest, and Support Vector Machine algorithm.

In Naive Bayes, a total of 102 features sets were explored resulting only 700 of 1000 features used in the training process. The performance of the model trained in the Naive Bayes is shown in Table 8. Column “AR” represents the average of recall; column “AP” indicated the average of precision; column “AF” shows the average score for F1; column “ACC” gives the accuracy score for the whole classes; column “ERR” illustrates the classification error rate; column ‘STD” indicates the value of the accuracy standard deviation. In Naıve Bayes model, Joy achieved the best F1 score (0,823), while Sadness received the lowest F1 score (0,798). The model achieved 0,806 for both F1 score and accuracy, with a standard deviation of accuracy around ±1, 3%. In the next training phase, a total of 102 features sets were explored in Generalized Linear Model, resulting in the best 650 features to be used in this phase. In the Generalized Linear Model, Anger achieved the best F1 score (0,940), while Fear received the lowest F1 score (0,858).

Table 4. Fast Large Margin Hyper-parameter Exploration

C	Classification Error
0.001	0.153
0.010	0.146
0.100	0.128
1	0.137
10	0.158
100	0.195
1000	0.219

Table 5. Decision Tree Hyper-parameter Exploration

Max Depth	Classification Error
2	0.4871
4	0.4724
7	0.4590
10	0.4451
15	0.4215
25	0.3931

This model archived the best precision (0,902), recall (0,902), accuracy (0,902), F1 score (0,901), and classification error rate (0,098). The standard deviation of the accuracy of this model is approximately ±1, 2%.

Table 6. Random Forest Hyper-parameter Exploration

Number of Trees	Max Depth	Classification Error
20	2	0.408
60	2	0.414
100	2	0.402
140	2	0.392
20	4	0.467
60	4	0.448
100	4	0.426
140	4	0.404
20	7	0.531
60	7	0.453
100	7	0.412
140	7	0,380

In Fast-Large Margin, a total of 102 features sets were explored, resulting only the best 650 of 1000 features used in the training process. This model also explores eight hyper-parameter settings of cost parameter, *C* (see table 4. The parameter indicates the penalty parameter of the error term. The parameter *C* was evaluated in 8 settings from 0, 001 to 1000 multiplied by 10 in each phase. The best result achieved by the parameter *C* of 0, 1 with classification error of 0, 128. Joy achieved the best F1 score (0,9), while Fear received the lowest F1 score (0,827). The model achieved 0,871 and 0,872 for F1 score and accuracy, respectively. This model achieved the best score for a standard deviation of the accuracy of approximately ±0, 3%. Next phase of training, Artificial Neural Network architecture was applied (see Table 3 for the hyper-parameters and architecture settings). With Artificial Neural Network, Anger achieved the best F1 score (0,913), while Fear received the lowest F1 score (0,849). The model achieved 0,884 and 0,885 for F1 score and accuracy, respectively, with a standard deviation of accuracy around ±0, 6%.

In Decision Tree, a total of 116 features sets were explored, resulting in only the best 550 of 1000 features used in the training process. This model also explores six hyper-parameter settings of maximum depth of the tree.

Table 7. Support Vector Machine Hyper-parameter Exploration

Gamma	C	Classification Error
0.005	10	0.187
0.050	10	0.188
0.500	10	0.190
5.000	10	0.205

0.005	100	0.188
0.050	100	0.188
0.500	100	0.152
5.000	100	0.205
0.005	1000	0.159
0.050	1000	0.155
0.500	1000	0.191
5.000	1000	0.212

demonstrates the settings of max depth with its classification error rate. The best result achieved by a maximum depth of 25 with a classification error of 0.393. Joy achieved the best F1 score (0.664), while Sadness received the lowest F1 score (0.504). The model archived the lowest performances compare to the others (0.587 and 0.608 for F1 score and accuracy, respectively). Next phase of training, a total of 315 features sets were explored in the Random Forest algorithm, resulting in the best 600 features to be used in this phase. In Random Forest Model, we also explored twelve combinations of the number of trees and the maximum depth of each tree in the forest. The number of trees ranged from 20-140, with a maximum depth of 2-7. Table 6 demonstrates the combination of both parameters. Joy achieved the best F1 score (0.696), while Fear received the lowest F1 score (0.411). The model achieved 0.601 and 0.62 for F1 score and accuracy, respectively. Both the Decision Tree and Random Forest are not the best algorithm to solve the problem. Both models resulted in a very high recall and extremely low precision in Anger (1 and 0,471 respectively for Decision Tree, 1 and 0.461 respectively for Random Forest), very high precision but extremely low recall in Fear (0.369 and 1 respectively for Decision Tree, 0.259 and 1 respectively for Random Forest, and Sadness (0.337 and 1 respectively for Decision Tree, 0.498 and 1 respectively for Random Forest). From the results, it can be concluded that both models tried to learn and fit the model as specific as possible to the training data, where the maximum number of trees and depth were chosen as the best outcome. However, both models can not generally represent the data in the testing data. Finally, In Support Vector Machine, a total of 1449 features sets were explored, resulting in only the best 250 of 1000 features used in the training process. This model also explores twelve hyper-parameter settings of γ and cost parameter, C (see table 7. The best result achieved by the setting of γ of 0.5 and C of 100 with a classification error of 0.152. Joy achieved the best F1 score (0.878), while Fear received the lowest F1 score (0.783).

5. CONCLUSION AND FUTURE WORK.

Several numbers of machine learning algorithms were implemented to train text-based emotions recognition on social media conversation. The algorithms are Naive Bayes, Generalized Linear Model, Fast-Large Margin, Artificial Neural Network, Decision Tree, Random Forest, and Support Vector Machine. This research aims to comprehensively explore the machine learning algorithms to provide the best model of text-based emotions recognition. The feature used in each algorithm depends on the optimal combination of the features sets with the lowest complexity and lowest classification error rates. The overall results are described in table 8, where Generalized Linear Model achieved the best precision (0.902), recall (0.90), accuracy (0.902), F1 score (0.901), and classification error rate (0.098). The standard deviation of the accuracy of this model is approximately ±1.2%. The second-best performances were achieved by Fast-Large Margin algorithm with a precision of 0.874, recall of 0.872, the accuracy of 0.872, F1 score of 0.871), and the classification error rate of 0.128. The model provides the best standard deviation of the accuracy of this model of approximately ±0.3%.

In contrast, the lowest performance archived by Decision Tree model with the precision of 0.772, recall of 0.607, the accuracy of 0.608, F1 score of 0.587), and the classification error rate of 0.392. The standard deviation of the accuracy of this model is approximately ±1.1%.

Table 8. Results

Algorithm	Class	Recall	Precision	F1	AR	AP	AF	ACC	ERR	STD
Naive Bayes	Anger	0.847	0.761	0.801	0.809	0.806	0.806	0.806	0.194	±1.3%
	Fear	0.753	0.857	0.802						
	Joy	0.804	0.844	0.823						
	Sadness	0.820	0.777	0.798						
GLM	Anger	0.976	0.905	0.940	0.902	0.902	0.901	0.902	0.098	±1.2%
	Fear	0.820	0.901	0.858						
	Joy	0.918	0.914	0.916						
	Sadness	0.894	0.887	0.891						
	Anger	0.969	0.834	0.897						

FLM	Fear	0.776	0.884	0.827	0.872	0.874	0.871	0.872	0.128	±0.3%
	Joy	0.878	0.922	0.900						
	Sadness	0.863	0.856	0.859						
ANN	Anger	0.973	0.861	0.913	0.885	0.886	0.884	0.885	0.115	±0.6%
	Fear	0.809	0.893	0.849						
	Joy	0.890	0.923	0.906						
	Sadness	0.867	0.867	0.867						
Decision Tree	Anger	1.000	0.471	0.641	0.607	0.772	0.587	0.608	0.392	±1.1%
	Fear	0.369	1.000	0.539						
	Joy	0.722	0.615	0.664						
	Sadness	0.337	1.000	0.504						
Random Forest	Anger	1.000	0.461	0.631	0.62	0.783	0.601	0.62	0.38	±0.5%
	Fear	0.259	1.000	0.411						
	Joy	0.722	0.672	0.696						
	Sadness	0.498	1.000	0.665						
SVM	Anger	0.875	0.838	0.856	0.848	0.85	0.849	0.85	0.15	±1.1%
	Fear	0,820	0,783	0,801						
	Joy	0.859	0.898	0.878						
	Sadness	0.839	0.881	0.859						

Decision Tree and Random Forest models tried to fit the models with the training data precisely; Hence, those models were not recommended in this problem. Overall, Anger and Joy consistently archived the best performance throughout the models, where Fear and Sadness are quite challenging to be recognized in some models. Several research ideas can be implemented as a future research direction. First, a local language (e.g. Indonesian) can be collected and trained using similar settings of research methods proposed in this research. Other deep learning algorithms, such as Long-Short Term Memory, Convolutional neural network, can be explored as the learning algorithm. Finally, the models can be implemented to the other application or affective systems such as virtual humans[1,18], and others.

REFERENCES

1. Chowanda, A., Blanchfield, P., Flintham, M., Valstar, M.. Erisa: Building emotionally realistic social game-agents companions. In: *International Conference on Intelligent Virtual Agents*. Springer; 2014, p. 134–143.
2. Zhu, W., Chowanda, A., Valstar, M.. Topic switch models for dialogue management in virtual humans. In: *International Conference on Intelligent Virtual Agents*. Springer; 2016, p. 407–411.
3. Hajar, M., et al. Using youtube comments for text-based emotion recognition. *Procedia Computer Science* 2016;83:292–299.
4. Bravo-Marquez, F., Frank, E., Pfahringer, B., Mohammad, S.M.. Affectivetweets: a weka package for analyzing affect in tweets. *Journal of Machine Learning Research* 2019;20:1–6.
5. Shivhare, S.N., Khethawat, S.. Emotion detection from text. *arXiv preprint arXiv:12054944* 2012;.
6. Sutoyo, R., Chowanda, A., Kurniati, A., Wongso, R.. Designing an emotionally realistic chatbot framework to enhance its believability with aiml and information states. *Procedia Computer Science* 2019;157:621–628.
7. Hajar, M., et al. Using youtube comments for text-based emotion recognition. *Procedia Computer Science* 2016;83:292–299.
8. Banik, N., Rahman, M.H.H.. Evaluation of Naıve bayes and support vector machines on bangla textual movie reviews. In: *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*. IEEE; 2018, p. 1–6.
9. Hasan, M., Rundensteiner, E., Agu, E.. Automatic emotion detection in text streams by analyzing twitter data. *International Journal of Data Science and Analytics* 2019;7(1):35–51.
10. Liu, Z.T., Wu, M., Cao, W.H., Mao, J.W., Xu, J.P., Tan, G.Z.. Speech emotion recognition based on feature selection and extreme learning machine decision tree. *Neurocomputing* 2018;273:271–280.
11. Zhang, X., Li, W., Ying, H., Li, F., Tang, S., Lu, S.. Emotion detection in online social networks: A multi-label learning approach. *IEEE Internet of Things Journal* 2020;.
12. Chowanda, A., Chowanda, A.D.. Recurrent neural network to deep learn conversation in indonesian. *Procedia computer science* 2017; 116:579–586.
13. Zheng, J., Zheng, L., Yang, L.. Research and analysis in fine-grained sentiment of film reviews based on deep learning. *Journal of Physics: Conference Series* 2019;1237(2):022152.
14. Wang, Z., Balasubramani, B.S., Cruz, I.F.. Predictive analytics using text classification for restaurant inspections. In: *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*. 2017, p. 1–4.

15. Ghiassi, M., Skinner, J., Zimbra, D.. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications* 2013;40(16):6266–6282.
16. Tausczik, Y.R., Pennebaker, J.W.. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology* 2010;29(1):24–54.
17. Jing, L.P., Huang, H.K., Shi, H.B.. Improved feature selection approach tfidf in text mining. In: *Proceedings. International Conference on Machine Learning and Cybernetics*; vol. 2. IEEE; 2002, p. 944–946.
18. Valstar, M., Baur, T., Cafaro, A., Ghitulescu, A., Potard, B., Wagner, J., et al. Ask alice: an artificial retrieval of information agent. In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. 2016, p. 419–420.