Frustration Level Analysis in Customer Support Tweets for Different Languages

Viktorija Ļeonova* and Jānis Zuters

University of Latvia, Faculty of Computing, Raina bulvaris 19, Riga, LV-1586, Latvia

Abstract
In this paper, we present comparative analysis of frustration intensity prediction for tweets in different languages using neural-network-driven models combining lexical and non-lexical means of expression. The different configurations of models were tested on customer support dialog texts in two languages – Latvian and English. The experimental results show the texts in both languages to be effectively evaluated for frustration intensity with slightly better overall results in Latvian. For both languages, the prediction models with configurations using all available features based on non-lexical means of expression yield the best accuracy, while the utilization of those features result in similar improvement in both languages.

Keywords
Machine learning, deep learning, neural networks, emotion annotation, frustration, non-lexical means of expression

1. Introduction

Living in a society requires a measure of one's relationship with the other. No coordination or collaboration is possible if an individual does not know what to expect from the other. For humans, as a social species, this is also true. Some scientists speculate [1] that our brain has developed because of our extensive social interactions for the purposes of navigating an ever-changing landscape in a closed group. And it is only natural that with the high noon of the Internet, especially Web 2.0 with its plenitude of user-generated content, the researchers would seek to try and formalize the recognition of emotions in digital media. The sheer volume of such media renders it nigh impossible for human processing, and that's when the automation comes into play.

Here, the same tremendous increase in processing power, storage volumes and interconnectivity that allowed the users to generate unsparing volumes of various media content, has provided for the development of technologies for harnessing those. And so, the researchers continuously sought to employ the most advanced techniques to annotate the emotions in user-generated content. Emotion recognition in general can serve a range of purposes, such as building a picture of a typical sentiment towards a public person or a phenomenon [2] or building an emotion-aware healthcare system [3].

In the very beginning, the emotion recognition mostly focused on speech, but as the social networks, such as Facebook and Twitter, gained more and more users [4] and voice communication relatively withered, the emotion recognition in text could not be ignored.

However, possibly due to the popularity of the base emotion system, proposed by Eckman [5] that discriminated between six basic emotions: anger, joy, sadness, surprise, disgust, and happiness, the most researchers concentrated on recognition of those, often amending the list by adding or removing one or the other. In fewer cases, the researchers employed a two- or three- dimensional model that assigns valence, arousal and dominance values to every emotion [6].

Baltic DB&IS 2022 Doctoral Consortium and Forum, July 03-06, 2022, Riga, LATVIA
*Corresponding author
EMAIL: viktorija.leonova@lu.lv (V. Leonova); janis.zuters@lu.lv (J. Zuters)
ORCID: 0000-0002-8070-8134 (V. Leonova); 0000-0002-3194-9142 (J. Zuters)

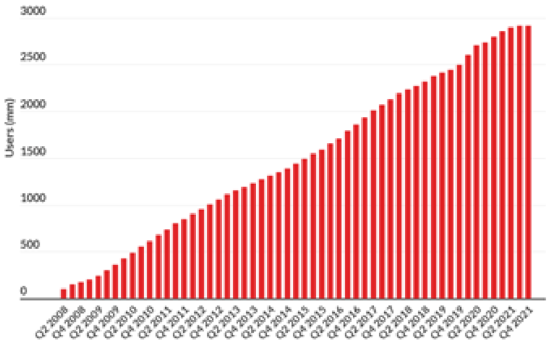

Figure 1: Millions of active Facebook users from 2008 to 2021. Source: [4]

This division, though, leaves unobserved such emotion as frustration. With virtually any company being present in major social networks and users contacting those on a daily basis, frustration, being a good measure of (dis)satisfaction with the company, may appear fairly helpful. However, nowadays there are only a handful of works that touch on the subject of frustration recognition, especially when talking about text. From the other facet, the existing projects, dedicated to emotion recognition in text, with a few exceptions concentrate only on the words and their sequences. They mainly ignore the fact that, unlike the published text and its descendants, Twitter and other social network messages contain a wide range of non-lexical entities, from omnipresent emojis to various ASCII and Unicode arts, such as ¯_(ツ)_/¯. In addition, most of those works are targeting English, while low-resource languages still struggle.

In our previous work [7] we have tested our hypothesis regarding the addition of features based on non-lexical means of expression (NLME) and found that those indeed improve the frustration recognition accuracy on Latvian dataset.

In this work, we seek to demonstrate that the addition of NLME features derived from Latvian dataset to the frustration recognition model is comparably beneficial when applied to the English dataset and thus the model is effectively language-independent and can be used for frustration recognition in English. Naturally, the application of this model to other languages is subject to testing and is currently limited to the means of expression shared between users of alphabetic languages. Its extension to the hieroglyphic languages is subject to studying and deriving the NLME used by the bearers of respective culture.

This paper has the following structure: at first, we present related works in Section 2; in the next section, we describe the datasets used for the experimentation. Section 4 presents the experimental setup and is followed by Section 5 that describes the presented model. Section 6 discusses the performance of the model and is followed by a review of possible future work and a conclusion.

2. Related works

As we have already mentioned, the emotion annotation in text and media was a subject of keen interest for the last couple of decades. The system capable of emotion recognition in speech and synthesis of emotional content was presented as early as in 1999 [8] and the very next year the researchers have employed neural networks for the automation of this process [9]. For a time, the emotion recognition was concentrated on speech, as the textual content was not nearly as ubiquitous as it has become later and did not play a prominent role in defining the sentiments and disposition of the Internet population. Only five years later there started to appear the researches aiming to derive emotions from text, first in multi-modal settings [10], where emotions derived from the textual content of a speech would play a supplementary role. However, as the developments in the field progressed, the system of automatic emotion recognition based solely on the text, started to emerge [11]. While those earlier works were mostly keyword-based, as the time passed, deep learning methods have started to be used in hybrid model in combination with the classic statistical methods [12] or alone [13], the current state of art presuming to use neural-network based model in combination with extensive vocabularies, encompassing the statistical weights of different words for various emotions, as well as word- and character-based n-gram features [14].

When we speak about the emotions being annotated, however, especially when we speak about low-resource languages with a small number of annotated corpora available for training models and calculating statistics, it can be seen that most authors cling to annotating models based on Ekman's six basic emotions: joy, anger, happiness, sadness, surprise and disgust. They are sometimes used in their original form, for example, [15], or in modified way, by adding or removing emotions from the list, with popular options being the Plutchik's [16] extension of the basic emotions by adding anticipation and trust as counterparts of surprise and disgust, for example, in [17] or addition of neutral emotion, such as in [18]. Another popular variant is reducing the list of the basic emotions by removing disgust [19] or both disgust and surprise [20], with more exotic variations ranging from replacing surprise with fondness [21] to recognizing 12, 15 and more finely discriminated emotions [14]. Less represented, but still universally recognized is using two-[22] or three-factor [23] models, that represent each emotion in the space of continual dimensions of valence, arousal and (in three-factor models) dominance. As it can be seen, frustration is very rarely part of the deal, appearing in only a few works, like [24], and is generally being understudied despite being potentially beneficial for such fields like customer support.

As most of the state-of-art models employ impressive language-specific vocabularies and n-gram features for annotation, the most researchers focus on English, as the language with enormous number of available resources available, and there are only a few resources targeting low-resource languages, such as Latvian, as, for example, [25], or being language-independent, with tangentially relevant examples including language-independent sentiment analysis [26] or language-independent emotion recognition in speech [27].

Non-lexical means of expression (NLME), potentially universal within the circle of languages sharing the same cultural context, have been studied sparingly and mostly in a dislocated manner, one or the other appearing in emotion annotation models. For example, usage of exclamation and question marks were used as a predictor in [28] and the message length in [29], but to the best of our knowledge, no systematic attempts were made before ours [7].

3. Datasets

For our experiments we have effectively used two datasets, in English and in Latvian. The English one represents the subset of Kaggle Twitter messages dataset[2], annotated for levels of frustration. It consists of 400 dialogs with 843 annotated user turns. The Latvian dataset contains 283 dialogs with 688 annotated user turns. User messages in both datasets were annotated by three independent annotators, and the median value was used as a resulting grade in further experiments. Both English and Latvian datasets, along with the code, are available in GitHub[3] and were described in detail in [30] and [7], respectively.

4. Non-lexical means of expression

In our work we are using the following non-lexical means of expression for predicting the frustration level. We would like to mention that the set of NLME features used in the experimentations represent a subset of features that we have identified originally, that had at least a weak correlation with the annotated grade. While the full correlation table and selection process is described in [7], we will note a couple examples. The message length had the highest (positive) correlation of 0.44 with the annotated frustration level, while most features, such as number of emojis, had a weak correlation around ±0.1. The selected features are:

- Length of the message
- Number of exclamation marks in the message
- Number of question marks in the message
- Number of commas in the message

[2] https://www.kaggle.com/thoughtvector/customer-support-on-twitter
[3] https://github.com/Lynx1981/dfrustration/tree/master/LatvianTweets

- Number of dots in the message
- Number of quotes in the message
- Number of uppercase words of the length greater than four characters
- Number of positive emotions made up of typographical marks
- Number of negative emotions made up of typographical marks
- Presence of picture in the message
- Presence of built-in smileys indiscriminate of valence in the message
- Mentions of Customer Protection Bureau / other accounts in the message

Those features encode the non-lexical characteristics of the message and as such, construe the second part of the model input sequence, as described in Section 6, where the proposed model is discussed.

5. Experimental framework for frustration intensity prediction

In order to test our hypothesis and find the best meta-parameters, we have constructed the following setup: we have constructed a NN-based model that accepts as an input a number of features, derived from a user message addressed to a customer support representative and assigns a grade representing the predicted level of frustration. This grade is compared to the grade assigned to this message by the annotators. The code of the model is available on GitHub along with the datasets used for training.

In our study, we have explored how the performance of the model is affected by the different set of input features, as well as by preceding segmentation of the message and different parameters of NN itself. Neural networks (multi-layered perceptrons with one hidden layer) were used as the technique to build our models as we have already successfully used them in the previous experiments and the main focus of the research is on frustration level analysis rather than a concrete machine learning used.

The overall experimentation was divided into two phases:

- Preparational phase of selecting experimental configurations empirically by conducting a few sets of experiments:
  a. Establishing hyperparameters for neural networks,
  b. Selection of the input configurations of lexical data,
  c. Establishing the best-performing set of NLME features.
- Main phase of running experiments to obtain final results.

First, we have established the optimal number of hidden neurons in the model, running the experiment with model configuration including 32, 64, 96 and 128 neurons. The results were that the model runs the best with 64 hidden neurons for both English and Latvian; so all further experiments were conducted using this configuration.

The next was assessing the role of preprocessing in overall performance. For this purpose, the model was run with the maximal number of input parameters, with the model configuration including 64 hidden neurons. The conclusion was that the results for the English dataset were ever so unexpectedly fully consistent with the ones obtained for Latvian data: the segmentation improved the accuracy of predictions by slightly more than one percent.

After the hidden neurons count and the effect of segmentation were established, we turned to establishing the best-performing set of NLME features. To research those, we have run yet another series of experiments using various combinations of input features. To name the most prominent, we have used the single best feature, removal of underperforming features, and all features present, along with the bag-of-words feature that remained unchanged. We established that the results were consistent with ones obtained on the Latvian dataset.

Were able to conclude that the behavior of the features is generally consistent over different languages, and the greater part of the accuracy is due to the four best predictors, while the complete removal of the features that produce no visible improvement when used in isolation, is disadvantageous to the resulting performance and leads to the slight decrease in accuracy.

As the performance of the original model on the English dataset was established, we have clarified whether the model adjusted for language and culture universality would not have the advantage of accuracy. For example, we used the number of PTAC (Consumer Rights Protection Bureau (of Latvia) mentions, which is inapplicable for English-speaking users; it was tentatively replaced with the total mention count, and the letter “a” repetition was complemented by letter “o” repetition, as it was the only repeated letter in English dataset. The resulting slight increase in accuracy confirmed the soundness of this replacement.

Performance assessment was conducted by computing accuracy as a percent of correct frustration level predictions via leave-one-out cross-validation that consisted of training the model on all data except one entry and comparing the frustration in-tensity predicted for the one remaining (left out) entry, repeated for all the entries, averaged across fifteen runs. We use two points of reference: a neural model that only uses lexical features as input and the same model applied to the Latvian dataset.

6. Language-independent model to measure frustration level

It stands to reason that to provide high quality customer support and customer care, especially in combination with profiling and other knowledge acquisition techniques, as well as for the purposes of triage in a resource-critical circumstances, it would be highly beneficial to be able to automatically predict the level of frustration expressed in a customer’s message. It is especially so, if such a model is language-independent and thus can be used for low-resource languages, for which no extensive vocabularies with annotated emotions and n-grams exist. Here we demonstrate that the proposed model, by utilizing the interactive vocabulary building principles and (to a limited extent) universal features based on non-lexical means of expression, demonstrate comparable performance in measuring the level of frustration for English and Latvian messages, addressed to company customer support representatives via Twitter social network.

The model that we have developed is predicting the frustration level on the scale of 0 to 4, zero denoting the absence of frustration and 4 representing the utmost level of frustration. To be able to do so, the model is taking advantage of three distinct features: interactive vocabulary construction, utilizing NLME-based features and initial data processing.

Interactive vocabulary construction: as discussed in [30], the first part of the features used as a model input, is selected based on the lexical means of expression — namely, words. During the training phase, all words in the training set are appraised for their predictive potential, by calculating an average value of frustration and its standard deviation among the messages in which this word appeared. Fig. 2 gives the example of this vocabulary, constructed for the English subset.

The model that we have developed is predicting the frustration level on the scale of 0 to 4, zero denoting the absence of frustration and 4 representing the utmost level of frustration. To be able to do so, the model is taking advantage of three distinct features: interactive vocabulary construction, utilizing NLME-based features and initial data processing.

Interactive vocabulary construction: as discussed in [30], the first part of the features used as a model input, is selected based on the lexical means of expression — namely, words. During the training phase, all words in the training set are appraised for their predictive potential, by calculating an average value of frustration and its standard deviation among the messages in which this word appeared. Fig. 2 gives the example of this vocabulary, constructed for the English subset.

```
becky (7, 6) [6, 0, 0, 0, 0, 0, 1] 0.0 0.0
ech (5, 5) [0, 0, 0, 5, 0, 0, 0] 3.0 0.0
sal (7, 5) [0, 0, 0, 5, 0, 0, 2] 3.0 0.0
channel (6, 5) [0, 0, 5, 0, 0, 1, 0] 2.0 0.0
unacceptable (5, 5) [0, 0, 0, 5, 0, 0, 0] 3.0 0.0
oct (5, 5) [0, 0, 5, 0, 0, 0, 0] 2.0 0.0
walmart (5, 5) [5, 0, 0, 0, 0, 0, 0] 0.0 0.0
queue (5, 4) [0, 0, 0, 4, 0, 1, 0] 3.0 0.0
playlists (5, 4) [0, 4, 0, 0, 0, 0, 1] 1.0 0.0
friends (5, 4) [0, 0, 4, 0, 0, 0, 1] 2.0 0.0
```

Figure 2: Top ten entries from the interactively constructed vocabulary. The numbers represent: 1) in round brackets: total number of messages, number of usable messages 2) In square brackets: number

of messages for each value of frustration level: 0 through 4, message incomprehensible, impossible to establish the level of frustration, 3) after round brackets: average value, standard deviation.

One hundred (as to why, see [30]) of the best predictor words — the ones with the lowest standard deviation — are used to create a bag-of-words. This means that the lexical part of every message is coded as a sequence of one hundred binary values, where every binary value, 0 or 1, represents whether the corresponding word from the vocabulary was present in the message. Figure 3 illustrates this process.

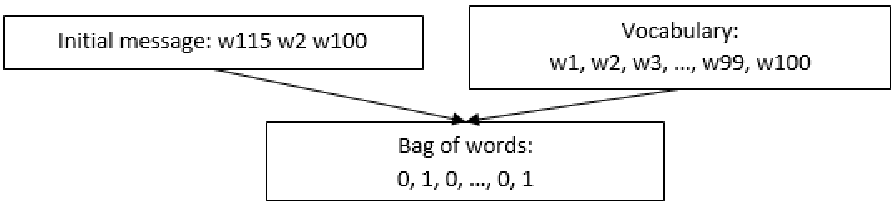

Figure 3: The first part (bag of words) model input construction.

The second part of the input is constructed using NLME features, described in Section 4: Non-Lexical Means of Expression. Along with the Bag-of-Words they form the model input.

The last important feature that we ought to mention is input processing. Before constructing the dictionary, the input message is processed by subword segmentation (using GenSeg tool [31]), which helps to alleviate the noise resulting from different grammatical forms being used in the same context. We would like to mention that it does not affect the dictionary construction; independently of segmentation usage, the vocabulary is built based on the available lexemes.

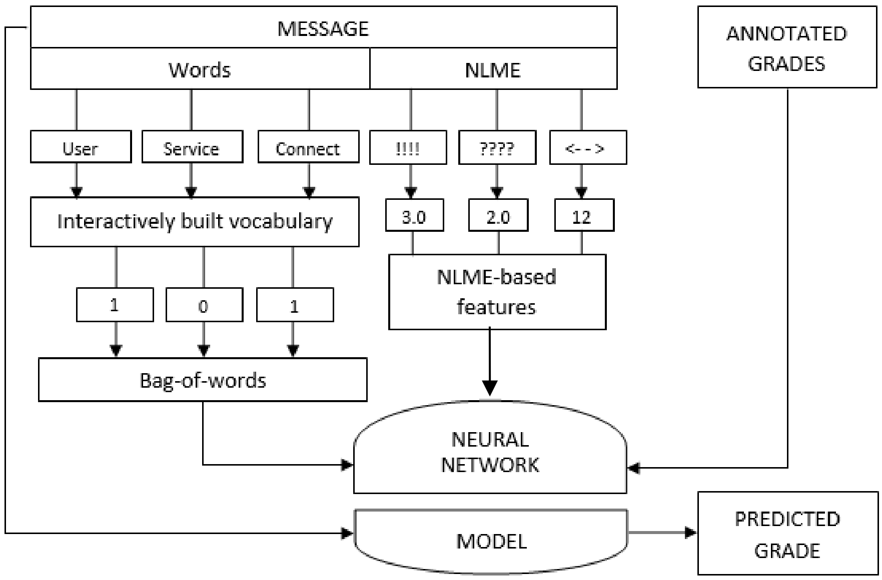

Figure 4: Frustration level predicting model

The detailed analysis of the experiment results is summarized in the following section.

7. Comparison of English and Latvian in frustration intensity prediction

For both languages, we have acquired a series of results for different model configurations, as described in section 4 “Experimental Framework”. Specifically, the following questions were addressed:

- How the model augmented with NLME features perform in comparison with the reference model when applied to English dataset and would the performance be dependent on these features similarly to the model applied to Latvian dataset?

- Would the segmentation improve the results similarly for both languages?

First of all, we gladly report that the two languages in their colloquial aspect appear to be similar enough so that the almost identical set of features added to the model would improve its performance by 6pp from 41% to 47% accuracy which is slightly lower than 7pp accuracy (42% to 49%) improvement achieved for Latvian. In both cases we are using as a baseline a bag-of-words model.

Not fully unexpected, we have found that the best performers were preserved: the most improvement was due to the set of four best features. Those are the length of the message, the number of exclamation marks in the message, the number of the question marks in the message, and the number of dots in the message, accordingly. However, in the manner similar to Latvian, removing the underperforming features did not improve the result, but otherwise: without features, seemingly not contributing to the performance, the overall result was slightly lower — they jointly contributed around 0.5pp for both Latvian and English, totaling 48.3% for Latvian and 46.7% for English.

However, it has to be mentioned that two of the NLME features used for Latvian was country- and language- specific, as we have used as an indication a number of PTAC (Customer Rights Protection Bureau) mentions in a message, which obviously is unusable in case of English tweets, as well as the repetitions of letter “a”, not characteristic for English language; there, the repetition of “o”, is more typical. In order to adjust the model to suit the dataset in English, we replaced it with a feature that calculates the number of “@” symbols that are universally used for mentions in various social networks and added the repetition of both vowels. This has improved the accuracy by 0.6pp, resulting in a total of 48.2% (7.3pp of improvement compared with the baseline). Curiously, the universalization also served in a good way for Latvian, improving the accuracy score by 0.3pp. The results are summarized in Tables 1 and 2.

Table 1

Frustration prediction accuracies (%) for various proposed models. C1 - NLME model with all features, C1* – mentions added, C2 - NLME model without subpar features, C3 - NLME model with all features and no segmentation, C4 - NLME model with a single best feature, RM – reference model

Model	C1	C1*	C2	C3	C4	RM
Latvian	48.8	49.1	48.4	47.5	46.9	42.1
English	47.2	48.2	46.7	45.9	43.7	40.9

Table 2

Frustration prediction improvements (pp) against the reference model for various proposed models. C1 - NLME model with all features, C1* – mentions added, C2 - NLME model without subpar features, C3 - NLME model with all features and no segmentation, C4 - NLME model with a single best feature.

Model	C1	C1*	C2	C3	C4
Latvian	6.7	7.0	6.3	5.4	4.8
English	6.3	7.3	5.8	5.0	2.8

What has come as a surprise, though, was the role of segmentation in the overall result. While for Latvian, being a synthetical language with a lot of flexions and grammatical forms, the slight improvement of 1.25pp achieved by segmentation of the source data, was to a certain degree expected, the same result of 1.25pp achieved for mostly analytical English was not.

Summarizing our findings, we can tell that the performance improvement resulting from extending the model with NLME features and data segmentation appear to be transferable to another language, namely English, from Latvian, for which this set of features was initially developed, to the full extent. That is, the extension of the bag-of-words model improves the results by 6pp or 7pp, of which the increase of 1.25pp is achieved due to the subword segmentation of the data. The removal of underperforming features causes the decline in resulting accuracy. The best results are acquired using 64 hidden neurons over 100 epochs.

8. Future works

In this study, we have researched whether the presumably language-independent model, originally developed using Latvian dataset, would be applicable to the English data. As we have demonstrated, it is indeed working as intended. However, Latvian and English both belong to the Indo-European language family, thus raising a question: do the applicability of the proposed model crosses the border of the language family, and whether, being sufficiently augmented, it could be applied to non-alphabetic languages. In the future, we would like to explore those possibilities. In addition, we want to research the possible extensions of the NLME set, should this prove possible.

9. Conclusion

The development of social networks and explosive growth of user-generated content made it nearly impossible to keep afloat without employing social media to keep in contact with the target audience, both providing the information and receiving feedback. Companies nowadays routinely use social networks to launch web-oriented campaigns and react to users' mentions and messages, however, due to the enormous volume of content it might be beneficial to employ one or the other automation technique in order to stay informed of relevant trends, tendencies, and sentiments. Emotion annotation plays a vital part in such methods and systems and thus keeps being the object of keen interest of numerous modern researchers. The existing works are mostly focusing on annotating basic emotions, while frustration is underrepresented despite being of practical interest in such areas as customer support, customer satisfaction and alike.

In our previous works, we have presented a neural network-based model that targeted measuring the level of frustration on the scale of 1 to 5 in the Twitter messages with interactively built vocabulary [30]. and showed how non-lexical means of expression and segmentation can improve the predictions [7] on the material of the annotated dataset in Latvian.

In this work, we observed the performance of the model developed on the material of Latvian dataset and the role of input segmentation when applied to the English dataset. For those purposes we have used datasets consisting of user dialogues with customer support representatives in Twitter and manually annotated. We have demonstrated that input data processing as well as the features initially developed on the Latvian material are providing a similar increase in accuracy, even more so after the small feature adjustment for higher extent of language-independence. As a baseline for comparison, we are using the accuracy, achieved by the model without the employment of data processing methods or NLME-based features. The baseline is approximately 42% for Latvian and 41% for English. The model, employing both NLME and data processing, achieves the accuracy of 47% for English and 49% for Latvian, which give 6pp and 7pp of increase in accuracy, respectively. However, provided the features are adjusted in accordance with English data, the resulting accuracy achieved on the English dataset comprises 48%, which is 7pp over the reference model.

10. References

[1] Whiten, A. and van de Waal, E., 2017. Social learning, culture and the ‘socio-cultural brain’ of human and non-human primates. Neuroscience & Biobehavioral Reviews, 82, pp.58-75.

[2] Wang, S., Schraagen, M., Sang, E.T.K. and Dastani, M., 2020. Public sentiment on governmental COVID-19 measures in Dutch social media.

[3] Ayata, D., Yaslan, Y. and Kamasak, M.E., 2020. Emotion recognition from multimodal physiological signals for emotion aware healthcare systems. *Journal of Medical and Biological Engineering*, *40*(2), pp.149-157.

[4] https://www.businessofapps.com/data/facebook-statistics/ as of 2022-02-17

[5] Ekman P.: An Argument for Basic Emotions. Cognition and Emotion, vol. 6(3-4), pp. 169–200. (1992)

[6] Mehrabian, A.: Basic Dimensions for A General Psychological Theory. pp. 39–53. (1980).
[7] Leonova, V. and Zuters, J., 2021, September. Frustration Level Annotation in Latvian Tweets with Non-Lexical Means of Expression. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)* (pp. 814-823).
[8] Moriyama, T. and Ozawa, S., 1999, June. Emotion recognition and synthesis system on speech. In *Proceedings IEEE International Conference on Multimedia Computing and Systems* (Vol. 1, pp. 840-844). IEEE.
[9] Nicholson, J., Takahashi, K. and Nakatsu, R., 2000. Emotion recognition in speech using neural networks. *Neural computing & applications*, *9*(4), pp.290-296.
[10] Chuang, Z.J. and Wu, C.H., 2004, August. Multi-modal emotion recognition from speech and text. In International Journal of Computational Linguistics & Chinese Language Processing, Volume 9, Number 2, August 2004: Special Issue on New Trends of Speech and Language Processing (pp. 45-62).
[11] Huang, X., Yang, Y. and Zhou, C., 2005, October. Emotional metaphors for emotion recognition in Chinese text. In *International Conference on Affective Computing and Intelligent Interaction* (pp. 319-325). Springer, Berlin, Heidelberg.
[12] Seol, Y.S., Kim, D.J. and Kim, H.W., 2008, July. Emotion recognition from text using knowledge-based ANN. In *ITC-CSCC: International Technical Conference on Circuits Systems, Computers and Communications* (pp. 1569-1572).
[13] Ghazi, D., Inkpen, D. and Szpakowicz, S., 2010, May. Hierarchical approach to emotion recognition and classification in texts. In *Canadian Conference on Artificial Intelligence* (pp. 40-50). Springer, Berlin, Heidelberg.
[14] Ameer, I., Sidorov, G., Gómez-Adorno, H. and Nawab, R.M.A., 2022. Multi-label Emotion Classification on Code-Mixed Text: Data and Methods. *IEEE Access*.
[15] Haryadi, D. and Kusuma, G.P., 2019. Emotion detection in text using nested long short-term memory. *11480 (IJACSA) International Journal of Advanced Computer Science and Applications*, *10*(6).
[16] Plutchik R. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4):344–350, 2001.
[17] Semeraro, A., Vilella, S. and Ruffo, G., 2021. PyPlutchik: Visualising and comparing emotion-annotated corpora. *Plos one*, *16*(9), p.e0256503.
[18] Feng, S., Wei, J., Wang, D., Yang, X., Yang, Z., Zhang, Y. and Yu, G., 2021. SINN: A speaker influence aware neural network model for emotion detection in conversations. *World Wide Web*, *24*(6), pp.2019-2048.
[19] Araque, O., Gatti, L., Staiano, J. and Guerini, M.: Depechemood++: A Bilingual Emotion Lexicon Built Through Simple Yet Powerful Techniques. *IEEE transactions on affective computing*. (2019)
[20] Mohammad, S., Bravo-Marquez, F., Salameh, M. and Kiritchenko, S., 2018, June. Semeval-2018 task 1: Affect in tweets. In *Proceedings of the 12th international workshop on semantic evaluation* (pp. 1-17).
[21] Yao, Y., Wang, S., Xu, R., Liu, B., Gui, L., Lu, Q. and Wang, X., 2014. The construction of an emotion annotated corpus on microblog text. *Journal of Chinese Information Processing*, *28*(5), pp.83-91.
[22] Hofmann, J., Troiano, E. and Klinger, R., 2021. Emotion-aware, emotion-agnostic, or automatic: Corpus creation strategies to obtain cognitive event appraisal annotations. *arXiv preprint arXiv:2102.12858*.
[23] Saif M. Mohammad. 2018. Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*. Melbourne, Australia.
[24] Hu, Tianran, Anbang Xu, Zhe Liu, Quanzeng You, Yufan Guo, Vibha Sinha, Jiebo Luo, and Rama Akkiraju. "Touch your heart: A tone-aware chatbot for customer care on social media." In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1-12. 2018.

[25] Gruzitis, N., Nespore-Berzkalne, G. and Saulite, B., 2018, May. Creation of Latvian FrameNet based on universal dependencies. In *Proceedings of the International FrameNet Workshop (IFNW)* (pp. 23-27).

[26] Shakeel, M.H., Faizullah, S., Alghamidi, T. and Khan, I., 2020, February. Language independent sentiment analysis. In *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)* (pp. 1-5). IEEE.

[27] Singh, R., Puri, H., Aggarwal, N. and Gupta, V., 2020. An efficient language-independent acoustic emotion classification system. *Arabian Journal for Science and Engineering*, *45*(4), pp.3111-3121.

[28] Kirk R., Roach M.A., Johnson J., Guthrie J., and Harabagiu S.M. "EmpaTweet: Annotating and Detecting Emotions on Twitter." In Lrec, vol. 12, pp. 3806-3813. 2012.

[29] Hautasaari, Ari, Naomi Yamashita, and Ge Gao. "How non-native English speakers perceive the emotional valence of messages in text-based computer-mediated communication." Discourse Processes 56, no. 1 (2019): 24-40.

[30] Zuters, J. and Leonova, V., 2020. Adaptive Vocabulary Construction for Frustration Intensity Modelling in Customer Support Dialog Texts. *International Journal of Computer Science & Information Technology (IJCSIT) Vol*, *12*.

[31] Zuters, J., Gus Strazds, and Leonova, V. and Viktorija Ļeonova. "Morphology-Inspired Word Segmentation for Neural Machine Translation." In *Databases and Information Systems X*, pp. 225-239. IOS Press, 2019.