

Extreme Learning Machines For Efficient Speech Emotion Estimation In Julia

Georgios Drakopoulos*, Phivos Mylonas

Department of Informatics, Ionian University, Tsirigoti Sq. 7, Kerkyra 49100, Hellas

Abstract

Speech is a mainstay of communication across literally all human activities. Besides facts and statements speech carries substantial information regarding experiences, thoughts, and emotions, therefore adding significant context. Moreover, non-linguistic elements such as pauses add more to the message. The field of speech emotion recognition (SER) has been developed precisely to develop algorithms and tools performing what humans learn to do from early on. One promising line of research comes from applying deep learning techniques trained on numerous audio attributes to discern between various emotions as dictated by a given model of fundamental human emotions. Extreme learning machines (ELMs) are neural network architectures achieving efficiency through simplicity and can potentially operate akin to a sparse coder. When trained by a plethora of audio attributes, such as cepstral coefficients, zero crossing rate, and autocorrelation, then it can classify emotions in speech based on the established emotion wheel model. The evaluation, done with the Toronto emotional speech set (TESS) on an ELM implemented in Julia, is quite encouraging.

Keywords

extreme learning machine, speech emotion recognition, emotion classification, Plutchik model, higher order patterns, spectrogram, cepstral coefficients, zero crossing rate, TESS dataset, Julia

1. Introduction

Language, whether oral or written, is among the major sources of human emotion and perhaps a mainstay of civilization itself. The field of speech emotion recognition (SER) almost since its formulation has been an essentially demanding field systematically garnering intense interdisciplinary interest since it aims to answer fundamental questions regarding human speech, which includes major elements such as intonation and pitch as well as latent and non-linguistic elements such as pauses and the length of sentences. Because of the complexity and volatility of human speech, SER relies heavily on machine learning (ML) and recently on deep learning (DL) techniques for performing tasks.

Human emotion models such as the *emotion wheel* by Plutchik [1] and the *universal emotion* models [2] have been developed to explain not only which emotions are fundamental, with interpretations ranging from social conditioning to brain functionality and evolutionary goals, but also how they are composed, which may well entail non-linear operations. In any case, such models can serve well as training guides to ML models for speech

emotion classification as is the case here.

Among the various models proposed for the various SER tasks, extreme learning machines (ELMs) have shown considerable potential. The latter can be partially at least attributed to the ELM structure which has only a single but very long hidden layer. In turn, this allows for straightforward and easy to interpret training schemes, all of which eventually stem from a synaptic weight regularization property. This is aligned with the intuition that a certain optimality condition should hold in order for the weights to be uniquely derived.

The primary research contribution of this conference paper is the development of an ELM implemented in Julia and operating like a sparse encoder for the emotion classification of sentences coming from the ubiquitous Toronto emotion speech set (TESS) collection, a benchmark for training ML and DL models for SER tasks.

The remainder of this work is structured as follows. The recent scientific literature regarding ELMs, SER, and graph mining is briefly reviewed in section 2. In section 3 the proposed methodology is described, whereas the results obtained using the TESS dataset are analysed in section 4. Possible future research directions are given in section 5. Bold capital letters denote matrices, bold small vectors, and small scalars. Acronyms are explained the first time they are encountered in the text. Finally, the notation of this work is summarized in table 1.

2. Related Work

Because of its interdisciplinary nature SER has been at the attention focus of a number of fields [3]. To address

CIKM'22: 31st ACM International Conference on Information and Knowledge Management (companion volume), October 17–21, 2022, Atlanta, GA

*Corresponding author.

✉ c16drak@ionio.gr (G. Drakopoulos); fmylonas@ionio.gr

(P. Mylonas)

🆔 0000-0002-0975-1877 (G. Drakopoulos); 0000-0002-6916-3129

(P. Mylonas)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

Table 1
Notation Summary

| Symbol | Meaning | First in |
|------------------------|-----------------------------|----------|
| $\stackrel{\Delta}{=}$ | Equality by definition | Eq. (1) |
| $\ \cdot\ $ | Vector or matrix norm | Eq. (6) |
| $\varphi(\cdot)$ | Activation function | Eq. (3) |
| $\tanh(\cdot)$ | Hyperbolic tangent function | Eq. (3) |
| $\text{tr}(\cdot)$ | Matrix trace | Eq. (7) |

the inherent complexity of the SER tasks, ML approaches such as ensemble learning [4], deep convolutional neural networks [5], domain invariant feature learning [6], two-dimensional convolutional neural networks [7], and multimodal deep learning [8] have been proposed in the scientific literature. Human emotion models such as the *emotion wheel* by Plutchik [1] or the *universal emotion theory* by Ekman [2] typically describe a fundamental set of emotions [9, 10, 11] along with composition rules and possible evolutionary explanations for them [12]. More recently, personality taxonomies go beyond single emotional reactions and treat personality as a whole such as the Myers-Brigs type indicator (MBTI). A reasoning based framework for emotion classification is [13].

ELMs have been used in ML because of the simplicity of their architecture [14]. They have been used as part of ML pipelines for wavelet transforms [15], in conjunction with an autoencoder for predicting the concentration of emitted greenhouse gases from boilers [16], and in optimizing a Kalman filter for determining the aging factors of supercapacitors [17]. Further applications include estimating soil thermal conductivity [18] and an evolving kernel assisted ELM for medical diagnosis [19], whereas an extensive list of applications is given in [20].

Graph mining is a field relying heavily on ML [21] and graph signal processing techniques [22]. Regarding the use of ML, self organizing maps (SOMs) for recommending cultural content are presented in [23], exploiting natural language attributes for finding linked requirements between software artefacts is the focus of [24], decompressing a sequence of Twitter graphs compressed with the two-dimensional discrete cosine transform using a tensor stack network (TSN) is described in [25], combining graph mining with transformers is shown in [26], advanced graph clustering techniques for classifying variation of cancer genomes [27], message passing graph neural networks for fuzzy [28] and ordinary Twitter graphs [29] are described, a GPU-based system for efficient graph mining is shown in [30], partitioning the user base of a portal for cultural content recommendation is explained in [31], visualizing massive graphs for human feedback is described in [32], approximating directed graphs with undirected ones under optimality conditions is shown in [33], classification of noisy graphs

is given in [34], sequential graph collaborative filtering is the topic of [35], mining hot sports in trajectories with graph based methodologies is developed in [36], and fMRI image classification with tensor distance metrics is the focus of [37].

3. Methodology

3.1. Attributes

Emotion models have been developed in order to explain how emotions work, their intensity and elicit conditions, how they may be composed in case of emotion levels, and possibly their evolutionary purpose. In this set of models the one proposed by Plutchik has been among the earliest and one of the most commonly used in engineering applications. Additionally, it has an easy to understand and intuitive-friendly visual interpretation, which is shown in figure 1. Notice that this figure depicts a two dimensional projection of a cone.

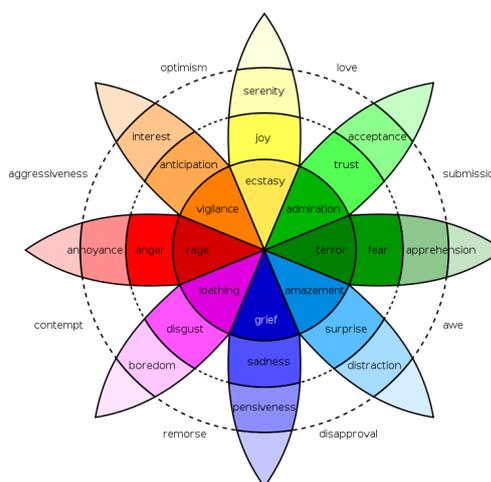


Figure 1: Plutchik model (From Wikipedia).

According to this model each emotion corresponds to a location in a circle which is primarily a function of its valence as well as of its direction. The latter is related to the nature of the emotion under consideration, which also determines at least in part its polarity. Specifically, there are in total eight directions with three scales each. Moreover, there are some emotions which are combinations of others from two directions. Moreover, the set of emotions is categorized as basic, primary, secondary, and tertiary. Primary emotions are archetypes the remaining ones are patterned after or are derived of. They are characterized by especially high survival value.

Table 2
Primary Emotions In Plutchik’s Model

| Emotion | Polarity | Opposite |
|--------------|----------------------|--------------|
| Neutral | Neutral | Neutral |
| Surprise | Positive or negative | Anticipation |
| Anticipation | Positive or negative | Surprise |
| Joy | Positive | Sadness |
| Trust | Positive | Disgust |
| Anger | Negative | Fear |
| Sadness | Negative | Joy |
| Disgust | Negative | Trust |
| Fear | Negative | Anger |

As stated earlier, each of the above emotions has an associated emotional polarity. For most emotions this polarity is clear, although in some cases such as surprise this has to be determined by the context. The intensity of each emotion essentially determines its location on a given affective axis. The higher the intensity, the more emotional a person is at a given time.

The primary emotions in the model of Plutchik are shown in table 2. Their location on the intensity scale and their relationship to other emotions are shown in figure 1. Moreover, the primary emotions come in four bipolar opposite emotions in the sense that they accomplish opposite objectives and their physical manifestations are considerably different. Bipolarity does not necessarily mean that in every pair there is one feeling with positive polarity, though this may be the case as in the pair of joy and sadness. Instead, both emotions in the anger and fear pair are both perceived as negative, but they are diametric opposites in the context of *fight or flight*.

3.2. Training

ELM training has a simpler compared to other neural network architectures since it has only one hidden layer with a large number p of processing neurons. With the proper training, each neuron can be specialized in a particular subset of the training set, which comprises of n data vectors. In this case the ELM output matrix \mathbf{H} has the elementwise structure of equation (1).

From its structure the i -th row of \mathbf{H} contains the output of each neuron for i -th data point $0 \leq i \leq n-1$, while the j -th column consists of the output of the j -th neuron $0 \leq j \leq p-1$ across all the available data points, preserving the order in which they were given to the ELM.

$$\mathbf{H} \triangleq \begin{bmatrix} h_0(\mathbf{x}_0) & h_1(\mathbf{x}_0) & \dots & h_{p-1}(\mathbf{x}_0) \\ h_0(\mathbf{x}_1) & h_k(\mathbf{x}_1) & \dots & h_{p-1}(\mathbf{x}_1) \\ \vdots & \vdots & \ddots & \vdots \\ h_0(\mathbf{x}_{n-1}) & h_1(\mathbf{x}_{n-1}) & \dots & h_{p-1}(\mathbf{x}_{n-1}) \end{bmatrix} \in \mathbb{R}^{n \times p} \quad (1)$$

The individual output of the j -th neuron can be computed from the nonlinear combination of equation (2). Therein q is the number of input neurons which is much smaller than that of the hidden neurons p , namely $q \ll p$, and equal to the dimensionality of each data point.

$$h_j(\mathbf{x}_i) \triangleq \varphi_j \left(\sum_{k=1}^q w_{j,k} \mathbf{x}_i[k] \right) = \varphi_j(\mathbf{w}_j^T \mathbf{x}_i) \quad (2)$$

The nonlinear activation function $\varphi_k(\cdot)$ may take a number of forms such as the logistic function or polynomial kernels. In this case it is the hyperbolic tangent function of (3). It has the advantage of being differentiable and of being the Bayes estimator of a bipolar source under additive white Gaussian (AWGN) noise.

$$\varphi_k(x; \beta_0) \triangleq \tanh(x; \beta_0) \quad (3)$$

The first derivative ψ of φ can be expressed as a second order polynomial of the latter as shown in (4). This expression is that of Malthus population models.

$$\psi(x; \beta_0) \triangleq \frac{\partial \varphi(x; \beta_0)}{\partial x} = \beta_0(1 - \varphi^2(x; \beta_0)) \quad (4)$$

The column synaptic weight vector \mathbf{w}_j is formed by stacking the q weights $w_{j,k}$ connecting the j -th hidden neuron with the k -input one. Moreover, this is also the j -th column of the synaptic weight matrix \mathbf{W} . If the data points are stacked on top of each other, then the input matrix \mathbf{X} is formed. Thus \mathbf{H} of (1) can be rewritten as in (5), where the function of (3) is elementwise applied.

$$\mathbf{H} \triangleq \varphi(\mathbf{W}\mathbf{X}^T) \quad (5)$$

In general ELMs, depending on their training formulation, can perform regularized least squares fitting in order to determine the optimal weights as in (6) where η_0 is a hyperparameter. Therein the regularization term adds robustness to the algorithmic minimization process. To this end, the nonlinear least squares problem of (6) was formulated, where the Frobenius matrix norm is used since it is differentiable. Also \mathbf{Y} is the ground truth matrix containing the one hot encoding of the eight primary emotions and \mathbf{W}^* is the solution.

$$\mathbf{W}^* \triangleq \operatorname{argmin} [J] = \operatorname{argmin} \left[\eta_0 \|\mathbf{W}\|_F^2 + \|\varphi(\mathbf{W}\mathbf{X}^T) - \mathbf{Y}\|_F^2 \right] \quad (6)$$

Expanding (6) and taking into consideration the expansion of the Frobenius norm the objective function J to be minimized can be recast as in (7). Because of the form Frobenius norm and that of the nonlinear activation function J not only is differentiable but it also has a single global minimum. Additionally, the regularization term ensures that synaptic weight sparsity also taken into consideration. Thus, minimizing J translates into finding the

weight set achieving a tradeoff between fitting the ELM response to the target response with the least possible energy. The latter can be considered as the explanation closest to that dictated by Occam's razor.

$$J = \eta_0 \text{tr}(\mathbf{W}^T \mathbf{W}) + \text{tr} \left((\varphi(\mathbf{W}\mathbf{X}^T) - \mathbf{Y})^T (\varphi(\mathbf{W}\mathbf{X}^T) - \mathbf{Y}) \right) \quad (7)$$

The minimization problem of (7) is a regularized nonlinear least squares problem. The hyperparameter η_0 determines the relative weight of the synaptic weight matrix sparsity compared to how well the ELM response matches the target response. The problem of (7) can be solved by a plethora of methodologies including iterative ones such as fixed point methods. However, they should take into consideration the nonlinear term introduced by the activation function. This can be accomplished by utilizing methods such as the Gauss-Newton or a regularized version thereof. In this work the *steepest descent* iterative method was selected because of its simplicity and because of the single global minimum J has, since the latter is essentially a sum of squares.

Furthermore, it can be argued that the proposed ELM, if properly trained, operates like a sparse coder with each activation neuron corresponding to a single emotion. This approach is clear it can be extended to an arbitrary number of emotions, provided of course that the appropriate attributes are available. However, the ELM proposed here can in fact discover the emotional direction and not the valence itself.

3.3. Attributes

In this subsection the various features used to train the ELM described here, their primary properties, and their respective meaning are explained. Said attributes are also shown in table 3 along with a brief explanation.

The cepstral coefficients $c[k]$ express a modified short term power spectrum of a signal $s[k]$ consisting of speech samples. They are derived by an algorithmic process which involves the following steps:

- The sequence is pre-emphasized such that higher frequencies receive an energy boost.
- The spectrum is smoothed with a window, usually a Hamming window of odd length.
- The power spectrum is translated in the nonlinear Mel scale where resolution is not constant.
- The logarithm of said power spectrum, which is always real, is computed.
- The coefficients of the inverse Fourier spectrum are the cepstral coefficients.

The natural meaning of the cepstral coefficients is that they represent a power spectrum where each frequency band has a resolution roughly inversely proportional to

its central frequency. This allows the details of a speech signal to be more discernible.

The spectrogram of a signal is a function of time and frequency and shows how its frequency content evolves in small time steps. Typically it can be obtained by the wavelet transform, by the short time Fourier transform (STFT), or by a bank of bandpass filters such as Gabor and shifted Chebyshev filters. In any case, the resulting heatmap has been transformed to a long column vector, which incurs some information loss as the spatial structure is lost. This is attributed to the fact that the proposed ELM is trained with data points with are real vectors. An architecture natively handling matrices may be more adept in this scenario.

The k -th autocorrelation coefficient $a[k]$ of any real-valued stationary sequence $s[i]$ is defined as the expected value of the sequence multiplied by a shifted version of itself by k positions. In practice these stochastic coefficients are often approximated by the sample mean of equation (8) under the assumption of ergodicity. Autocorrelation coefficients are a measure of the self-similarity of the sequence under consideration and play a central role in discovering higher order patterns through the Wiener filter. It should be noted that the higher k is, the less reliable the estimation of $a[k]$ becomes as fewer term pairs are available. Therefore, k in most engineering applications is small compared to the total length n of the speech sample sequence. As a direct consequence of the Cauchy-Schwarz inequality, the maximum autocorrelation coefficient is the first one $a[0]$.

$$a[k] \approx \frac{1}{n-k} \sum_{i=0}^{n-k-1} s[i]s[i+k], \quad 0 \leq k \leq n-1 \quad (8)$$

Finally, the zero crossing rate (ZCR) is an important feature which assumes that the mean value of the speech signals has been subtracted from it during a preprocessing phase. ZCR is closely tied with the primary mode of the Hilbert-Huang spectrum (HHS), which is built on fundamental signals inherent in the sequence. Thus, intuitively speaking, the HHS is very similar to the Fourier spectrum but it is composed of basis signals progressively extracted from the original signal itself and hence having irregular shapes instead of weighted complex exponentials. In this context ZCR plays a role analogous to that of the fundamental frequency in Fourier analysis.

The audio attributes used in this work are also shown in table 3 along with their interpretation.

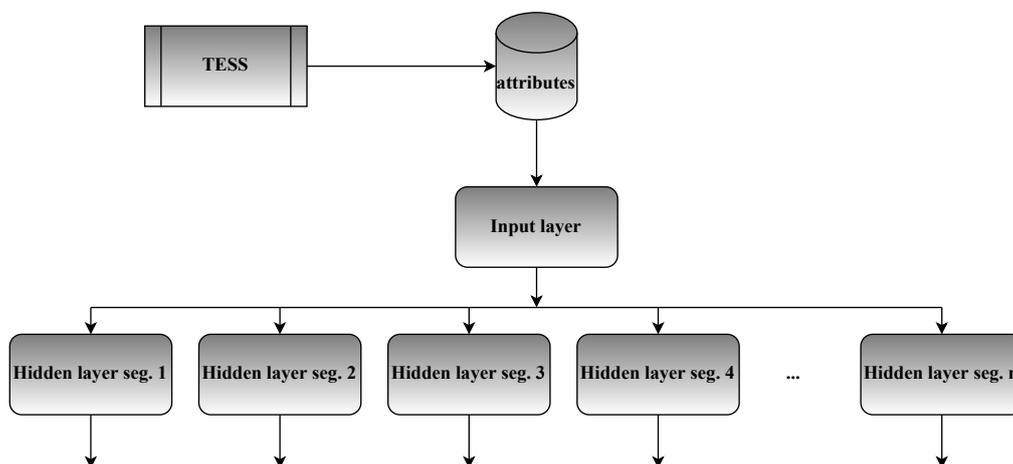
4. Results

4.1. ELM Architecture

The architecture of the proposed ELM is shown in 2. Notice that all hidden neurons belong to the same ELM layer

Table 3
Audio Attributes

| Attribute | Meaning |
|-----------------------|---|
| Cepstral coefficients | Short term windowed power spectrum |
| Spectrogram | Frequency content evolution over short time steps |
| Autocorrelation | Self-similarity patterns in the speech sequence |
| Zero crossing rate | Tied to primary mode of Hilbert-Huang spectrum |

**Figure 2:** Proposed architecture.

and they are conceptually but not physically segmented to show they are an integer multiple of the neurons of the input layer. Hence the hidden layer can be thought of as comprising of segments, although in practice all hidden neurons are simultaneously trained.

The implementation language of choice was Julia. It is a rapidly emerging multiparadigm high level language aiming at computation-heavy tasks such as those frequently encountered in DL and ML scenarios, large database clustering, extensive and fine grained simulations, and graph signal processing.

4.2. Emotion Recognition

The TESS dataset contains 200 target words spoken in the context of a carrier phrase by two actresses, a younger and an older one aged 26 and 64 respectively. Each recording contains 2000 data points, which are sufficient for processing, and they represent the neutral state plus six of the primary emotions according to Plutchik's model, namely these of anger, disgust, fear, happiness, pleasant surprise, and sadness. Therefore, from the emotions listed in table 2 anticipation and trust are absent. Consequently, from the four pairs of primary bipolar emotions only two are fully present in TESS.

In figure 3 is shown the heatmap resulting from the analysis of the ELM training. From it the following can be immediately inferred:

- The neutral emotional state is the only one which can be accurately discovered in the context of this work. This can be attributed to the fact that compared to the other states there is no valence. In turn this allows its isolation from the rest of the states in the attribute space with a margin sufficient for the ELM to discern it.
- On the contrary anger is the most difficult to be discovered. A possible explanation is that its bipolar opposite emotion is also present in TESS and, thus, certain instances have been misattributed to it. Moreover, anger is also confused with surprise and sadness. The former is possibly due to valence, whereas the latter because of polarity.
- Concerning the other bipolar pair of sadness and happiness, they are clearly distinguished from each other, but nevertheless there is a small probability they will be misclassified respectively as disgust and as pleasant surprise. This can be attributed to their valence as well as to the semantics of each emotion under consideration.
- The remaining emotions can be also be distin-

guished relatively easy from the others in the dataset. Still, the negative emotions tend to be classified with a lower level accuracy compared to the positive ones, with the single exception of sadness. This can be explained by their prevalence in TESS.

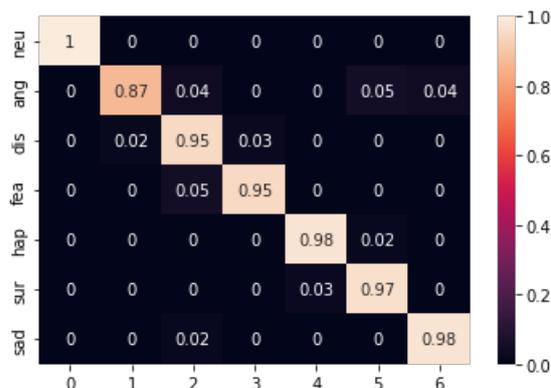


Figure 3: ELM heatmap.

In summary, the heatmap reveals a performance level which may be satisfactory for certain applications. Still, as negative emotions with the sole exception of sadness tend to be less accurately identified compared to the positive ones, there is room for improvement.

5. Conclusions

The focus of this conference paper is the development of an extreme learning machine (ELM) for speech emotion recognition (SER) based on the primary emotions identified in Plutchik’s model. Based on a wide array of audio attributes an ELM is trained to act like a sparse coder with the nine fundamental emotions are one-hot encoded in an output vector. The proposed approach is flexible enough as the training phase on an ELM is much simpler compared to that of other neural network architectures, especially the fundamental multilayer perceptron. The results obtained with waveforms taken from the established Toronto emotional speech set (TESS) are very encouraging in terms of accuracy.

Regarding future research directions, the proposed neural network architecture and the associated encoding can be tested with other publicly available speech datasets such as the Emo-Soundscape or SUSAS. Moreover, an ELM can be adapted to other human emotion models such as the *big five* or the *universal emotion theory*. Finally, attribute vectorization can be avoided with archi-

tectures capable of natively handling two-dimensional attributes such as the class of graph neural networks.

Acknowledgments

This conference paper is part of Project 451, a long term research initiative with a primary objective of developing novel, scalable, numerically stable, and interpretable higher order analytics.

References

- [1] A. Semeraro, S. Vilella, G. Ruffo, PyPlutchik: Visualising and comparing emotion-annotated corpora, *PLoS one* 16 (2021).
- [2] A. Talipu, A. Generosi, M. Mengoni, L. Giraldi, Evaluation of deep convolutional neural network architectures for emotion recognition in the wild, in: *ISCT, IEEE*, 2019, pp. 25–27.
- [3] T. M. Wani, T. S. Gunawan, S. A. A. Qadri, M. Kartiwi, E. Ambikairajah, A comprehensive review of speech emotion recognition systems, *IEEE Access* 9 (2021) 47795–47814.
- [4] W. Zehra, A. R. Javed, Z. Jalil, H. U. Khan, T. R. Gadekallu, Cross corpus multi-lingual speech emotion recognition using ensemble learning, *Complex & Intelligent Systems* 7 (2021) 1845–1854.
- [5] S. Kwon, Optimal feature selection based speech emotion recognition using two-stream deep convolutional neural network, *International Journal of Intelligent Systems* 36 (2021) 5116–5135.
- [6] C. Lu, Y. Zong, W. Zheng, Y. Li, C. Tang, B. W. Schuller, Domain invariant feature learning for speaker-independent speech emotion recognition, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30 (2022) 2217–2230.
- [7] Z. Zhao, Q. Li, Z. Zhang, N. Cummins, H. Wang, J. Tao, B. W. Schuller, Combining a parallel 2D CNN with a self-attention dilated residual network for CTC-based discrete speech emotion recognition, *Neural Networks* 141 (2021) 52–60.
- [8] S. Zhang, X. Tao, Y. Chuang, X. Zhao, Learning deep multimodal affective features for spontaneous speech emotion recognition, *Speech Communication* 127 (2021) 73–81.
- [9] P. Sreeja, G. Mahalakshmi, Emotion models: A review, *International Journal of Control Theory and Applications* 10 (2017) 651–657.
- [10] K. R. Scherer, et al., Psychological models of emotion, *The neuropsychology of emotion* 137 (2000) 137–162.
- [11] S. Marsella, J. Gratch, P. Petta, et al., Computational models of emotion, *A Blueprint for Affective*

- Computing - A sourcebook and manual 11 (2010) 21-46.
- [12] R. M. Nesse, Evolutionary explanations of emotions, *Human nature* 1 (1990) 261-289.
- [13] A. Lieto, G. L. Pozzato, S. Zoia, V. Patti, R. Damiano, A commonsense reasoning framework for explanatory emotion attribution, generation and re-classification, *Knowledge-Based Systems* 227 (2021).
- [14] Q.-Y. Zhu, A. K. Qin, P. N. Suganthan, G.-B. Huang, Evolutionary extreme learning machine, *Pattern recognition* 38 (2005) 1759-1763.
- [15] S. Yahia, S. Said, M. Zaied, Wavelet extreme learning machine and deep learning for data classification, *Neurocomputing* 470 (2022) 280-289.
- [16] Z. Tang, S. Wang, X. Chai, S. Cao, T. Ouyang, Y. Li, Auto-encoder-extreme learning machine model for boiler NOx emission concentration prediction, *Energy* 256 (2022).
- [17] D. Li, S. Li, S. Zhang, J. Sun, L. Wang, K. Wang, Aging state prediction for supercapacitors based on heuristic Kalman filter optimization extreme learning machine, *Energy* 250 (2022).
- [18] N. Kardani, A. Bardhan, P. Samui, M. Nazem, A. Zhou, D. J. Armaghani, A novel technique based on the improved firefly algorithm coupled with extreme learning machine (ELM-IFF) for predicting the thermal conductivity of soil, *Engineering with Computers* 38 (2022) 3321-3340.
- [19] J. Xia, D. Yang, H. Zhou, Y. Chen, H. Zhang, T. Liu, A. A. Heidari, H. Chen, Z. Pan, Evolving kernel extreme learning machine for medical diagnosis via a disperse foraging sine cosine algorithm, *Computers in Biology and Medicine* 141 (2022).
- [20] S. Ding, X. Xu, R. Nie, Extreme learning machine and its applications, *NCAA* 25 (2014) 549-556.
- [21] M. A. Thafar, S. Albaradie, R. S. Olayan, H. Ashoor, M. Essack, V. B. Bajic, Computational drug-target interaction prediction based on graph embedding and graph mining, in: *Proceedings of the 2020 10th international conference on bioscience, biochemistry and bioinformatics*, 2020, pp. 14-21.
- [22] K. Yamada, Y. Tanaka, Temporal multiresolution graph learning, *IEEE Access* 9 (2021) 143734-143745.
- [23] G. Drakopoulos, I. Giannoukou, S. Sioutas, P. Mylonas, Self organizing maps for cultural content delivery, *NCAA* (2022). doi:10.1007/s00521-022-07376-1.
- [24] M. Singh, Using natural language processing and graph mining to explore inter-related requirements in software artefacts, *ACM SIGSOFT Software Engineering Notes* 44 (2022) 37-42.
- [25] G. Drakopoulos, E. Kafeza, P. Mylonas, L. Iliadis, Transform-based graph topology similarity metrics, *NCAA* 33 (2021) 16363-16375. doi:10.1007/s00521-021-06235-9.
- [26] I. Tyagin, A. Kulshrestha, J. Sybrandt, K. Matta, M. Shtutman, I. Safro, Accelerating COVID-19 research with graph mining and transformer-based learning, in: *Conference on Artificial Intelligence*, volume 36, AAAI, 2022, pp. 12673-12679.
- [27] G. Gomez-Sanchez, L. Delgado-Serrano, D. Carrera, D. Torrents, J. L. Berral, Author correction: Clustering and graph mining techniques for classification of complex structural variations in cancer genomes, *Scientific Reports* 12 (2022).
- [28] G. Drakopoulos, E. Kafeza, P. Mylonas, S. Sioutas, A graph neural network for fuzzy Twitter graphs, in: G. Cong, M. Ramanath (Eds.), *CIKM companion volume*, volume 3052, CEUR-WS.org, 2021.
- [29] G. Drakopoulos, I. Giannoukou, P. Mylonas, S. Sioutas, A graph neural network for assessing the affective coherence of Twitter graphs, in: *IEEE Big Data, IEEE*, 2020, pp. 3618-3627. doi:10.1109/BigData50022.2020.9378492.
- [30] L. Hu, L. Zou, A GPU-based graph pattern mining system, in: *CIKM*, 2022, pp. 4867-4871.
- [31] G. Drakopoulos, Y. Voutos, P. Mylonas, S. Sioutas, Motivating item annotations in cultural portals with UI/UX based on behavioral economics, in: *IISA, IEEE*, 2021. doi:10.1109/IISA52424.2021.9555569.
- [32] S. A. Bhavsar, V. H. Patil, A. H. Patil, Graph partitioning and visualization in graph mining: A survey, *Multimedia Tools and Applications* (2022) 1-42.
- [33] G. Drakopoulos, I. Giannoukou, P. Mylonas, S. Sioutas, On tensor distances for self organizing maps: Clustering cognitive tasks, in: *DEXA*, volume 12392 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 195-210. doi:10.1007/978-3-030-59051-2_13.
- [34] Z. Xu, B. Du, H. Tong, Graph sanitation with application to node classification, in: *Web Conference*, ACM, 2022, pp. 1136-1147.
- [35] Z. Sun, B. Wu, Y. Wang, Y. Ye, Sequential graph collaborative filtering, *Information Sciences* 592 (2022) 244-260.
- [36] S. Wang, X. Niu, P. Fournier-Viger, D. Zhou, F. Min, A graph based approach for mining significant places in trajectory data, *Information Sciences* 609 (2022) 172-194.
- [37] G. Drakopoulos, E. Kafeza, P. Mylonas, S. Sioutas, Approximate high dimensional graph mining with matrix polar factorization: A Twitter application, in: *IEEE Big Data, IEEE*, 2021, pp. 4441-4449. doi:10.1109/BigData52589.2021.9671926.