Understanding the Role of Affect Dimensions in Detecting Emotions from Tweets: A Multi-task Approach

Rajdeep Mukherjee
IIT Kharagpur, India
rajdeep1989@iitkgp.ac.in

Atharva Naik*
IIT Kharagpur, India
atharvanaik2018@iitkgp.ac.in

Sriyash Poddar*
IIT Kharagpur, India
poddarsriyash@iitkgp.ac.in

Soham Dasgupta
MAIS Bangalore, India
sohamdasgupta91@gmail.com

Niloy Ganguly
IIT Kharagpur, India
niloy@cse.iitkgp.ac.in

ABSTRACT

We propose **VADEC**, a multi-task framework that exploits the correlation between the *categorical* and *dimensional* models of emotion representation for better subjectivity analysis. Focusing primarily on the effective detection of emotions from tweets, we jointly train *multi-label emotion classification* and *multi-dimensional emotion regression*, thereby utilizing the inter-relatedness between the tasks. Co-training especially helps in improving the performance of the *classification* task as we outperform the strongest baselines with 3.4%, 11%, and 3.9% gains in *Jaccard Accuracy*, *Macro-F1*, and *Micro-F1* scores respectively on the *AIT* dataset [17]. We also achieve state-of-the-art results with 11.3% gains averaged over six different metrics on the *SenWave* dataset [27]. For the *regression* task, *VADEC*, when trained with *SenWave*, achieves 7.6% and 16.5% gains in *Pearson Correlation* scores over the current state-of-the-art on the *EMOBANK* dataset [5] for the Valence (V) and Dominance (D) affect dimensions respectively. We conclude our work with a case study on COVID-19 tweets posted by Indians that further helps in establishing the efficacy of our proposed solution.

CCS CONCEPTS

• **Information systems → Sentiment analysis**.

KEYWORDS

Coarse-grained Emotion Analysis; Fine-grained Emotion Analysis; Valence-Arousal-Dominance; Multi-task Learning; Twitter; COVID

ACM Reference Format:
Rajdeep Mukherjee, Atharva Naik, Sriyash Poddar, Soham Dasgupta, and Niloy Ganguly. 2021. Understanding the Role of Affect Dimensions in Detecting Emotions from Tweets: A Multi-task Approach. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3404835.3463080

*Both authors contributed equally to this research.

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

ACM ISBN 978-1-4503-8037-9/21/07…$15.00
https://doi.org/10.1145/3404835.3463080

1 INTRODUCTION

With the proliferation of social media, as more and more people express their opinions online, detecting human emotions from their written narratives, especially tweets has become a crucial task given its widespread applications in e-commerce, public health monitoring, disaster management, etc. [17, 18]. *Categorical* models of emotion representation such as Plutchik's *Wheel of Emotion* [21] or Ekman's *Basic Emotions* [8] classify affective states into discrete categories (joy, anger, etc.). *Dimensional* models on the other hand describe emotions relative to their fundamental dimensions. Russel and Mehrabian's *VAD* model [23] for instance interprets emotions as points in a 3-D space with *Valence* (degree of pleasure or displeasure), *Arousal* (degree of calmness or excitement), and *Dominance* (degree of authority or submission) being the three orthogonal dimensions. Accordingly, the literature on text-based emotion analysis can be broadly divided into *coarse-grained classification* systems [10, 12–14, 28] and *fine-grained regression* systems [22, 24, 29, 30]. Although a *coarse-grained* approach is better-suited for the task of detecting emotions from tweets as observed in [4], prior works fail to exploit the direct correlation between the two models of emotion representation for finer interpretation. We utilize the better representational power of *dimensional* models [4] to improve the *emotion classification* performance by proposing **VADEC** that jointly trains *multi-label emotion classification* and *multi-dimensional emotion regression* in a multi-task framework.

Multi-task learning [6] has been successfully used across a wide spectrum of NLP tasks including emotion analysis [1, 30]. While *AAN* [30] takes an adversarial approach to learn discriminative features between two emotion dimensions at a time, *All_In_One* [1] proposes a multi-task ensemble framework to learn different configurations of tasks related to coarse- and fine-grained sentiment and emotion analysis. However, none of the methods combine the supervisions from VAD and categorical labels. Our proposed framework (Section 2) consists of a **classifier** module that is trained for the task of multi-label emotion classification, and a **regressor** module that co-trains the regression tasks corresponding to the V, A, and D dimensions. Owing to the unavailability of a common annotated corpus, the two tasks are trained using supervisions from their respective benchmark datasets (reported in Section 3.1), which further justifies the utility of our proposed multi-task approach.

VADEC learns better shared representations by jointly training the two modules, that especially help in improving the performance of the *classification* task, thereby achieving state-of-the-art results on the *AIT* [17] and *SenWave* [27] datasets (Section 3.3). For the

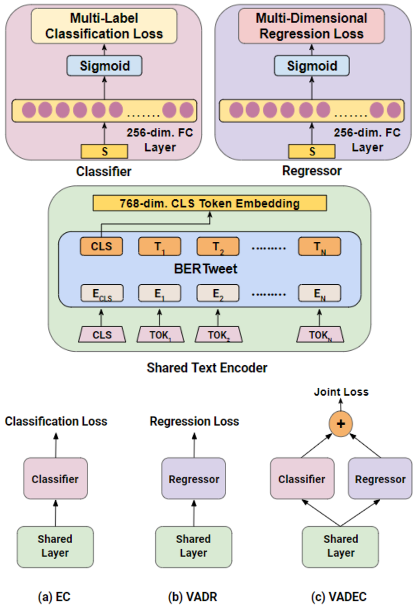


Figure 1: Components and Model Architecture: Pre-trained BERTweet serves as the *Shared Text Encoder* between the *Classifier* and *Regressor* modules.(a) *EC* and (b) *VADR* respectively represent the Multi-label Emotion *Classifier* and Multi-dimensional Emotion *Regressor* when trained individually. (c) *VADEC* represents our Multi-Task Affect Classifier that co-trains the two modules by optimizing the joint loss.

regression task, we achieve SOTA results on the *EMOBANK* dataset [5] for *V* and *D* dimensions (Section 3.4). We conclude our work with a detailed case study in Section 3.5, where we apply our trained multi-task model to detect and analyze the changing dynamics of Indian emotions towards the COVID-19 pandemic from their tweets. We discover the major factors contributing towards the various emotions and find their trends to correlate with real-life events.

2 *VADEC* ARCHITECTURE

Figure 1 illustrates the architecture of *VADEC*, that jointly trains a multi-label emotion *classifier* and a multi-dimensional emotion *regressor* with supervision from their respective datasets. Since we primarily focus on detecting emotions from tweets, we use *BERTweet* [19] to serve as our text-encoder. It is shared by the two modules and is hereby referred to as the *shared layer*. The 768-dim. [*CLS*] token embedding of the sentence/tweet obtained from BERTweet is first passed through a fully connected (FC) layer with 256 neurons in both the modules respectively. The *classifier* passes this intermediate representation through another FC layer with 11 output neurons, each activated using *Sigmoid* with a threshold of 0.5 to predict the presence/absence of one of the 11 emotion categories. *Binary Cross-Entropy* (BCE) with L2-norm regularization is used as the loss function, hereby referred to as the EC_{Loss}. Similarly, the *regressor* passes the 256-dim. intermediate representation through an FC layer with 3 output neurons (with *Sigmoid* activation) corresponding to the V, A and D dimensions. It then jointly optimizes the *Mean Squared Error* (MSE) loss of all three dimensions, hereby referred to as the $VADR_{Loss}$. *VADEC* jointly trains the two modules by optimizing the following multi-task objective:

$$VADEC_{\text{Loss}} = \lambda \cdot EC_{Loss} + (1 - \lambda) \cdot VADR_{Loss} \tag{1}$$

Here, λ represents a balancing parameter between the two losses. The weighted joint loss backpropagates through the *shared layer*, thereby fine-tuning the *BERTweet* parameters end-to-end.

3 RESULTS AND DISCUSSION

3.1 Datasets

For our experiments, we consider *EMOBANK*, a VAD dataset, and two categorical datasets, *AIT* and *SenWave* as described below:

- **EMOBANK** (Buechel and Hahn [5]) : A collection of around 10k English sentences from multiple genres (8,062 for training, and 1K sentences each for validation and testing), each annotated with continuous scores (in the range of 1 to 5) for *Valence*, *Arousal*, and *Dominance* dimensions of the text.
- **AIT** (Mohammad et al. [17]) : Created as part of SemEval 2018 Task 1: “Affect in Tweets”, it consists of 10,983 English tweets (6,838 for training, 886 for validation, 3,259 for testing), each with labels denoting the presence/absence of a total of 11 emotions.
- **SenWave** (Yang et al. [27]) : Till date the largest fine-grained annotated COVID-19 tweets dataset consisting of 10K English tweets (8K for training, and 1K each for validation and testing), each with corresponding labels denoting the presence/absence of 11 different emotions specific to COVID-19.

3.2 Experimental Setup

For all our model variants, we perform extensive experiments with different sets of hyper-parameters and select the best set w.r.t. lowest validation loss. Before evaluating the performance on the test set, we combine the training and validation data and re-train the models with the best obtained set of hyper-parameters (learning rate = $2e-5$, weight decay = 0.01, $\lambda = 0.5$, and no. of epochs = 5 for *VADEC*). For the *regression* task, the outputs of *Sigmoid* activation at each of the three output neurons are suitably scaled before calculating the *MSE* loss since the ground-truth VAD scores are in the range of 1-5. As **model ablations**, we investigate the role played by features derived from affect lexicons by additionally appending a 194-dim. *Empath*[1] [9] feature vector to the intermediate representations learnt by our model variants to be used for final predictions. Parameters of our *shared encoder* are initialized with pre-trained model weights (*roberta-base* for RoBERTa, and *bertweet-base* for BERTweet) from the *HuggingFace Transformers* library [25]. Other model parameters are randomly initialized. All our model variants are trained end-to-end with *AdamW* optimizer [16] on Tesla P100-PCIE (16GB) GPU. We additionally ensure the reproducibility of our results and make our code repository [2] publicly accessible.

3.3 Evaluating Emotion Classification

We first discuss the comparative results of our model variants and ablations on the **AIT** dataset. We then respectively report our state-of-the-art results achieved on the *AIT* and the **SenWave** datasets.

[1] https://github.com/Ejhfast/empath-client
[2] https://github.com/atharva-naik/VADEC

AIT Dataset

As **metrics** we use *Jaccard Accuracy*, *Macro-F1*, and *Micro-F1* [17]. Among recent **baselines**: (i) **BERTL** (Park et al. [20]) denotes the scores obtained by fine-tuning BERT-Large [7] on the *AIT* dataset, and (ii) **NTUA-SLP** (Baziotis et al. [3]) represents the winning entry for this (sub)task of SemEval 2018 Task 1 [17], where the authors take a transfer learning approach by first pre-training their Bi-LSTM architecture, equipped with multi-layer self attentions, on a large collection of general tweets and the dataset of SemEval 2017 Task 4A, before fine-tuning their model on this dataset. Among our **model variants and ablations**: (i) **EC** represents our *classifier* module, when trained as a single task (Fig. 1a), (ii) $\mathbf{EC_{RoBERTa}}$ uses *RoBERTa* [15] instead of *BERTweet* as the shared layer.

From Table 1, *NTUA-SLP* surprisingly outperforms *BERTL* (on *Jac. Acc.* and *Micro-F1*), a heavier model with 336M parameters. *EC* (trained with *BERTweet*) comfortably beats $EC_{RoBERTa}$ demonstrating the better efficacy of *BERTweet* in learning features from tweets. The sparse *Empath* feature vectors do not however add any value to the rich 768-dim. contextual representations learnt using BERT-based methods. We obtain our **best results with *VADEC***, with respectively **3.4%, and 3.9% gains in *Jacc. Acc.*, and *Micro-F1*** over *NTUA-SLP*, and **11% gain** in ***Macro-F1*** over *BERTL*.

SenWave Dataset

Considering the superior performance of *VADEC* over all its model variants and ablations from Table 1, here we directly compare the results of *VADEC*, re-trained with *SenWave* [27], with the ones reported by the authors of [27], serving as the only available **baseline** on this dataset. Following [27], we use *Label Ranking Average Precision* (LRAP), *Hamming Loss*, and *Weak Accuracy* (Accuracy) as **metrics** in addition to the ones reported in Table 1. As observed from Table 2, ***VADEC* achieves SOTA** by outperforming the baseline scores with **11.3%** performance gain averaged over **all 6 metrics**.

Overall, our results from Tables 1 and 2 demonstrate the advantage of utilizing the VAD supervisions for improving the performance of the multi-label emotion classification task.

3.4 Evaluating Emotion Regression

Pearson Correlation Coefficient r is used as the evaluation **metric** for this task. All the models are evaluated on the *EMOBANK* dataset. Among recent **baselines**: (i) **AAN** (Zhu et al. [30]) employs adversarial learning between two attention layers to learn discriminative word weight parameters for scoring two emotion dimensions at a time. The authors report the VAD scores for all 6 domains and 2 perspectives of *EMOBANK*. For comparison, we use their highest correlation score for each dimension, (ii) **All_In_One** (Akhtar et al. [1]) represents a multi-task ensemble framework which the authors use for learning four different configurations of multiple tasks related to emotion and sentiment analysis, (iii). **SVR-SLSTM** (Wu et al. [26]) represents a semi-supervised approach using variational autoencoders to predict the VAD scores, and (iv). **BERTL (EB ← AIT)** [20], the current state-of-the-art, fine-tunes BERT-Large [7] on the *AIT* dataset to predict VAD scores by means of minimizing EMD distances between the predicted VAD distributions and sorted categorical emotion distributions as a proxy for target VAD distributions. For comparison, we use their reported scores obtained upon further fine-tuning their best-trained model on the *EMOBANK* corpus. Our **model variants** include (i) **VADR** which represents our *regressor* module, when trained as a single task (Fig. 1b), (ii) $\mathbf{VAD_{RoBERTa}}$, an ablation where we experiment with *RoBERTa* as the shared layer, (iii) **VADEC (AIT)**, and (iv) **VADEC (SenWave)** representing the scores of our multi-task model when trained respectively with the *AIT* and *SenWave* datasets.

Table 1: Comparative Results on the *AIT*. Results of *VADEC* are statistically significant than *EC* with 95% conf. interval.

Methods	Jaccard Acc.	F1-Macro	F1-Micro
BERTL [20]	0.572	0.534	0.697
NTUA-SLP [3]	0.588	0.528	0.701
$EC_{RoBERTa}$	0.592	0.570	0.712
w/ Empath	0.585	0.562	0.706
EC	0.605	0.581	0.723
w/ Empath	0.602	0.570	0.720
VADEC	**0.608**	**0.593**	**0.728**
Significance T-Test (p-values)	0.029	-	-

Table 2: Comparative Results on the *SenWave* dataset.

Methods	Accuracy	Jac. Acc.	F1-Macro	F1-Micro	LRAP	Ham. Loss
SenWave [27]	0.847	0.495	0.517	0.573	0.745	0.153
VADEC	**0.877**	**0.560**	**0.563**	**0.620**	**0.818**	**0.123**

Table 3: Comparison of Pearson Correlation (r-values) for the emotion regression task on the *EMOBANK* (EB) dataset.

Methods	Valence (V)	Arousal (A)	Dominance (D)
AAN [30]	0.424	0.351	0.265
All_In_One [1]	0.635	0.375	0.277
SRV-SLSTM [26]	0.620	0.508	0.333
BERTL (EB ← AIT) [20]	0.765	**0.583**	0.416
$VADR_{RoBERTa}$	0.804	0.494	**0.511**
w/ Empath	0.798	0.482	0.510
VADR	0.821	0.553	0.493
VADEC (AIT)	0.820	0.563	0.459
VADEC (SenWave)	**0.823**	0.553	0.485

From Table 3, $VADR_{RoBERTa}$ shows the highest correlation (0.511) on the *D* dimension. *VADR* (w/ BERTweet) however outperforms $VADR_{RoBERTa}$ on the other two dimensions. Contrary to our observations in the *classification* task, co-training does not help in improving the performance of the *regression* task, as can be confirmed from the results of *VADEC* (AIT) and *VADR*. Although we are outclassed by *BERTL (EB ← AIT)* on the *A* dimension, *VADEC* (AIT) comfortably outperforms *BERTL (EB ← AIT)* on the *V* and *D* dimensions. **VADEC (SenWave)** further outclasses both *VADEC* (AIT) and *BERTL (EB ← AIT)* on *V* and *D* with **7.6% and 16.5% gains** respectively. To conclude, although joint-learning does not help the *regression* task as much as it helps in improving the *classification* performance (which in fact is our main objective), we still achieve noticeable improvements in majority of emotion dimensions.

3.5 COVID-19 and Indians: A Case Study

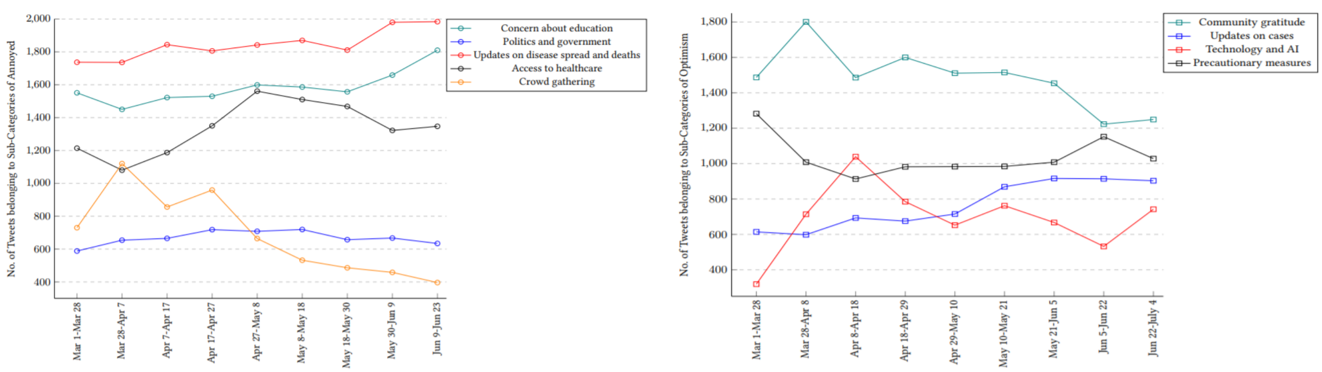
For this analysis, we consider **Twitter_IN**, a subset of *COVID-19 Twitter chatter* dataset (version 17) [2], containing around 140K English tweets from India posted between January 25th and July 4th 2020. Owing to very few reported cases in India before March 2020, we begin our analysis by predicting emotions from tweets, posted on or after Match 1st 2020, using VADEC trained on *EMOBANK*

Table 4: **Few Examples of Single and Multi-label Predictions on Tweets from *Twitter_IN***

Tweet	Predicted Labels
Single Label	
Let us spare a moment and thought for the junior resident doctors of Mumbai on the frontline fighting it out alone with little help from the government against all odds and at great personal risk	Thankful
This is the time to fight Covid19 at present but some intelligent Generals are focusing on war and terrorism	Annoyed
Multiple Labels	
The first Covid 19 positive from Meghalaya Dr John Sailo Rintathiang passed away early this morning. Sailo 69 who was also the owner of Bethany hospital was tested positive on April 13 2020	Sad, Official Report
Media is so obsessed with a particular community that they even misspell coronavirus	Annoyed, Joking, Surprise

Table 5: Major aspects affecting various emotions among Indians towards the COVID-19 pandemic.

Emotion	Major aspects
Annoyed	govt, politics, death, news, religion, jamaat, work, China, assault, border
Sad	lockdown, death, distancing, life, family, economy, village, doctor, worker, school
Thankful	doctor, service, staff, nurse, app, fund, assistance, leadership
Optimistic	initiative, opportunity, measure, arogyasetuapp, IndiaFightsCorona, stayhome, vaccine, change, support, action



(a) Annoyed **(b) Optimism**

Figure 2: Change in Sub-categories of Emotional Triggers towards the COVID-19 pandemic over time.

and *SenWave*. Few tweets with their predicted emotions are listed in Table 4. For each emotion, we obtain its contributing aspects by training an unsupervised neural topic model, ABAE (He et al. [11]) on the subset of tweets containing the given emotion as per *VADEC* predictions. Few emotions along with their most accurate aspects are reported in Table 5. For each emotion, the extracted aspect terms are further filtered and assigned meaningful sub-categories by means of a many-to-many mapping. In Figure 2, we plot the temporal trends of these sub-categories (with roughly equal-sized bins in terms of no. of tweets predicted with the emotion plotted) that respectively made Indians feel *annoyed* (Fig. 2a) and *optimistic* (Fig. 2b) over time. In Fig. 2a, the peak in *Crowd gathering* between March 28th and April 7th can be attributed to the *Tablighi Jamaat* gatherings[3] unfortunately triggering widespread criticism. Fig. 2b shows a high level of *Community gratitude* in general, with occasional peaks which may be attributed to the events targeted at raising solidarity among the public. For *Technology and AI*, we observe a peak near the launch date of the *Arogya Setu App*[4] - developed by the Indian Government to identify COVID-19 clusters.

[3] https://en.wikipedia.org/wiki/2020_Tablighi_Jamaat_COVID-19_hotspot_in_Delhi
[4] https://en.wikipedia.org/wiki/Aarogya_Setu

4 CONCLUSION AND FUTURE WORK

In this work, we for the first time exploit the correlation between *categorical* and *dimensional* models of emotion analysis by proposing *VADEC*, a multi-task affect classifier with the primary objective of efficiently detecting emotions from tweets. Co-training the tasks of *multi-label emotion classification* and *multi-dimensional emotion regression* helps the former thereby achieving state-of-the-art results on two benchmark datasets, *AIT* (non-COVID) and *SenWave* (COVID-related). For the *regression* task, *VADEC* still outperforms the strongest baseline on the *EMOBANK* dataset on the *V* and *D* dimensions. In future, we would like to investigate the hierarchical relationship between the tasks and analyze the relative impact of each emotion dimension on the emotion classification task.

ACKNOWLEDGMENTS

This research is partially supported by IMPRINT-2, a national initiative of the Ministry of Human Resource Development (MHRD), India. Niloy Ganguly was partially funded by the Federal Ministry of Education and Research (BMBF), Germany (grant no. 01DD20003).

REFERENCES

[1] S. Akhtar, D. Ghosal, A. Ekbal, P. Bhattacharyya, and S. Kurohashi. 2019. All-in-One: Emotion, Sentiment and Intensity Prediction using a Multi-task Ensemble Framework. IEEE Transactions on Affective Computing (2019), 1–1. https://doi.org/10.1109/TAFFC.2019.2926724

[2] Juan M. Banda, Ramya Tekumalla, Guanyu Wang, Jingyuan Yu, Tuo Liu, Yuning Ding, and Gerardo Chowell. 2020. A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration. ArXiv (2020). https://doi.org/10.5281/zenodo.3930903

[3] Christos Baziotis, Athanasiou Nikolaos, Alexandra Chronopoulou, Athanasia Kolovou, Georgios Paraskevopoulos, Nikolaos Ellinas, Shrikanth Narayanan, and Alexandros Potamianos. 2018. NTUA-SLP at SemEval-2018 Task 1: Predicting Affective Content in Tweets with Deep Attentive RNNs and Transfer Learning. In Proceedings of The 12th International Workshop on Semantic Evaluation. Association for Computational Linguistics, New Orleans, Louisiana, 245–255. https://doi.org/10.18653/v1/S18-1037

[4] Sven Buechel and Udo Hahn. 2016. Emotion Analysis as a Regression Problem — Dimensional Models and Their Implications on Emotion Representation and Metrical Evaluation. In Proceedings of the Twenty-Second European Conference on Artificial Intelligence (The Hague, The Netherlands) (ECAI'16). IOS Press, NLD, 1114–1122. https://doi.org/10.3233/978-1-61499-672-9-1114

[5] Sven Buechel and Udo Hahn. 2017. EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. Association for Computational Linguistics, Valencia, Spain, 578–585. https://www.aclweb.org/anthology/E17-2092

[6] R. Caruana. 2004. Multitask Learning. Machine Learning 28 (2004), 41–75.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[8] P. Ekman. 1992. An argument for basic emotions. Cognition & Emotion 6 (1992), 169–200. https://doi.org/10.1080/02699939208411068

[9] Ethan Fast, Binbin Chen, and Michael S. Bernstein. 2016. Empath: Understanding Topic Signals in Large-Scale Text. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 4647–4657. https://doi.org/10.1145/2858036.2858535

[10] Hao Fei, Yue Zhang, Yafeng Ren, and Donghong Ji. 2020. Latent Emotion Memory for Multi-Label Emotion Classification. Proceedings of the AAAI Conference on Artificial Intelligence 34, 05 (Apr. 2020), 7692–7699. https://doi.org/10.1609/aaai.v34i05.6271

[11] Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017. An Unsupervised Neural Attention Model for Aspect Extraction. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Vancouver, Canada, 388–397. https://doi.org/10.18653/v1/P17-1036

[12] Chenyang Huang, Amine Trabelsi, Xuebin Qin, Nawshad Farruque, and Osmar R. Zaïane. 2019. Seq2Emo for Multi-label Emotion Classification Based on Latent Variable Chains Transformation. arXiv:1911.02147 [cs.CL]

[13] Mohammed Jabreel and Antonio Moreno. 2019. A Deep Learning-Based Approach for Multi-Label Emotion Classification in Tweets. Applied Sciences 9, 6 (2019). https://doi.org/10.3390/app9061123

[14] Pratik Kayal, Mayank Singh, and Pawan Goyal. 2020. Weakly-Supervised Deep Learning for Domain Invariant Sentiment Classification. In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD (Hyderabad, India) (CoDS COMAD 2020). Association for Computing Machinery, New York, NY, USA, 239–243. https://doi.org/10.1145/3371158.3371194

[15] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692 [cs.CL] http://arxiv.org/abs/1907.11692

[16] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In International Conference on Learning Representations. https://openreview.net/forum?id=Bkg6RiCqY7

[17] Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. SemEval-2018 Task 1: Affect in Tweets. In Proceedings of The 12th International Workshop on Semantic Evaluation. Association for Computational Linguistics, New Orleans, Louisiana, 1–17. https://doi.org/10.18653/v1/S18-1001

[18] Saif Mohammad and Svetlana Kiritchenko. 2018. Understanding Emotions: A Dataset of Tweets to Study Interactions between Affect Categories. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Miyazaki, Japan. https://www.aclweb.org/anthology/L18-1030

[19] Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020. BERTweet: A pre-trained language model for English Tweets. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Association for Computational Linguistics, Online, 9–14. https://doi.org/10.18653/v1/2020.emnlp-demos.2

[20] Sungjoon Park, Jiseon Kim, Jaeyeol Jeon, Heeyoung Park, and Alice Oh. 2019. Toward Dimensional Emotion Detection from Categorical Emotion Annotations. CoRR abs/1911.02499 (2019). http://arxiv.org/abs/1911.02499

[21] ROBERT PLUTCHIK. 1980. Chapter 1 - A GENERAL PSYCHOEVOLUTIONARY THEORY OF EMOTION. In Theories of Emotion, Robert Plutchik and Henry Kellerman (Eds.). Academic Press, 3–33. https://doi.org/10.1016/B978-0-12-558701-3.50007-7

[22] Daniel Preoţiuc-Pietro, H. Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elisabeth Shulman. 2016. Modelling Valence and Arousal in Facebook posts. In Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Association for Computational Linguistics, San Diego, California, 9–15. https://doi.org/10.18653/v1/W16-0404

[23] James A. Russell and Albert Mehrabian. 1977. Evidence for a three-factor theory of emotions. Journal of Research in Personality 11, 3 (1977), 273 – 294. https://doi.org/10.1016/0092-6566(77)90037-X

[24] Jin Wang, Liang-Chih Yu, K. Robert Lai, and Xuejie Zhang. 2016. Dimensional Sentiment Analysis Using a Regional CNN-LSTM Model. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Association for Computational Linguistics, Berlin, Germany, 225–230. https://doi.org/10.18653/v1/P16-2037

[25] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Association for Computational Linguistics, Online, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

[26] Chuhan Wu, Fangzhao Wu, Sixing Wu, Zhigang Yuan, Junxin Liu, and Yongfeng Huang. 2019. Semi-supervised dimensional sentiment analysis with variational autoencoder. Knowledge-Based Systems 165 (2019), 30–39. https://doi.org/10.1016/j.knosys.2018.11.018

[27] Qiang Yang, Hind Alamro, Somayah Albaradei, Adil Salhi, Xiaoting Lv, Changsheng Ma, Manal Alshehri, Inji Jaber, Faroug Tifratene, Wei Wang, Takashi Gojobori, Carlos M. Duarte, Xin Gao, and Xiangliang Zhang. 2020. SenWave: Monitoring the Global Sentiments under the COVID-19 Pandemic. arXiv:2006.10842 [cs.SI]

[28] Jianfei Yu, Luís Marujo, Jing Jiang, Pradeep Karuturi, and William Brendel. 2018. Improving Multi-label Emotion Classification via Sentiment Classification with Dual Attention Transfer Network. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium, 1097–1102. https://doi.org/10.18653/v1/D18-1137

[29] Liang-Chih Yu, Jin Wang, K. Robert Lai, and Xue-jie Zhang. 2015. Predicting Valence-Arousal Ratings of Words Using a Weighted Graph Method. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Association for Computational Linguistics, Beijing, China, 788–793. https://doi.org/10.3115/v1/P15-2129

[30] Suyang Zhu, Shoushan Li, and Guodong Zhou. 2019. Adversarial Attention Modeling for Multi-dimensional Emotion Regression. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Florence, Italy, 471–480. https://doi.org/10.18653/v1/P19-1045