EMPIRICAL EMOTION RECOGNITION WITH LEVERAGING UNLABELED DATA USING ENHANCED COLLABORATIVE SEMI-SUPERVISED LEARNING

KANAGALA HARI KRISHNA, CHETTUPALLI BHAVANI

ASSISTANT PROFESSOR, DEPT OF IT, VIGNAN'S LARA INSTITUTE OF TECHNOLOGY AND SCIENCE, VADLAMUDI, ANDHRA PRADESH 522213.

MCA STUDENT, VIGNAN'S LARA INSTITUTE OF TECHNOLOGY AND SCIENCE, VADLAMUDI, ANDHRA PRADESH 522213.

Abstract:

One of the significant impediments that must be confronted while applying automatic emotion recognition to realistic human-machine interaction systems is the shortage of labeled data for training a strong model. Spurred by this worry, this paper tries to most extreme endeavor unlabeled data that are unavoidably accessible in reality and simple to be gathered, by methods for novel semi-supervised learning (SSL) approaches. Regular SSL strategies, for example, self-training, experience the ill effects of their inborn downside of error accumulation, i.e., the samples that are misclassified by the framework are persistently utilized to prepare the model in the accompanying learning iterations. To address this significant issue, we initially propose an enhanced learning strategy, by which we reexamine the already automatically labeled samples in each learning iteration, so as to refresh the training set by amending the mislabeled samples. We further endeavor multiple modalities and models in the SSL framework, by utilizing collaborative SSL, where all modalities and models are considered all the while; samples are selected by methods for limiting the joint entropy. This strategy should not just improve the exhibition of the model for data annotation and therefore upgrade the trustability of the automatically labeled data, yet in addition to hoist the assorted variety of selected data. To assess the viability of the proposed approaches, we performed broad investigations on the remote collaborative and full of feeling database, which incorporates multimodal accounts of unconstrained compelling interactions of dyads. The exact outcomes show that the proposed approaches altogether outflank as of late entrenched SSL techniques.

Index Terms—enhanced semi-supervised learning, collaborative learning, audiovisual emotion recognition.

I. INTRODUCTION

Automatic emotion recognition has pulled in wide consideration in artificial intelligence over the previous decade, since it assumes a fundamental job in accomplishing normal and agreeable human–machine interactions [1]–[2][3][4][5]. In any case, one significant hindrance that blocks its expansive applications, in actuality, settings is the absence of adequate labeled data as far as amount and decent variety, which is respected to be of high significance to fabricate a strong and proficient recognition model [6]–[7][8].

In view of the open accessibility of huge unlabeled data that can be effortlessly gathered through inescapable electronic gadgets [8], [9], one common arrangement comes to leveraging the estimation of these data in a successful manner. Semi-Supervised Learning (SSL) has been developed as a promising methodology since it intends to proficiently utilize machines (i.e., recognition models) to automatically 'comment on' unlabeled data, with (basically) no need of manual intercession. In the course of recent years, a few endeavors have been made and have indicated the advantages of SSL for emotion recognition.

In [10], Wu et al. presented a realistic based SSL model for emotion recognition from music, by which the oversight information (or the name data) is engendered from the labeled data to the unlabeled data by figuring the acoustic and label comparability among melodies. In [11], Schels et al. utilized a thickness estimation of every single accessible datum to move the mark data to unlabeled data. Comparative work was additionally announced in [12], however for the content based emotion characterization.

Rather than these transductive SSL approaches where both labeled and unlabeled data are considered to play out a forecast on the unlabeled data, more examination endeavors need to follow an inductive SSL worldview, fundamentally because of the way that the incredible ability of discriminative models (e.g., Neural Networks) for emotion recognition has been much of the time appeared over the previous decade [13]. In the inductive worldview, a predictive model is pre-constructed uniquely on the labeled data and afterward utilized for foreseeing the unlabeled data. For instance, Zhang et al. [14] utilized a normal inductive SSL approach called self-training to investigate the unlabeled data from various databases for emotion recognition from discourse. Moreover, co-training was proposed to abuse two perspectives (include sets) for emotion recognition. For instance, Zhang et al. [15] and [16] split the acoustic highlights into two gatherings (e.g., vitality or unearthly related), every one of which is viewed as one 'see' for emotion recognition from discourse. In like manner, Li et al. [17] took the individual and generic (i.e., the sentence whose subject isn't an individual) assessments as two 'sees' for emotion recognition from text. As of late, profound neural system based SSL has risen an extraordinary potential technique attributable to its capacity to distil significant level portrayals. Most as of late, Deng et al. [18] presented a mutual covered up layer structure with perform various tasks learning, which comprises of two undertakings – remaking inputs (autoencoder way) and anticipating emotions (grouping way). It is normal that the information can be moved from unlabeled data to labeled data through the autoencoder way.

Nonetheless, a large portion of these examinations just centered around a sign methodology, i.e., either sound [19], video [20], [21], or text [17]. These days, perceiving emotion by means of multiple modalities rises to be conspicuous , not just because of the wide utilization of cameras and receivers as previously mentioned, yet in addition because of the way that the blend of different modalities can regularly offer preferred execution over unimodality for emotion recognition .Nevertheless, multimodal data is frequently overlooked in many past SSL research. Unique in relation to past examinations, in this article we expect to utilize multiple modalities in SSL for emotion recognition.

Besides, conventional SSL approaches frequently experience the ill effects of an issue of execution debasement. That is, while adding all the more automatically commented on data to the training set frequently brings about more awful, as opposed to better, execution of recognition models [0]–[1][2]. To a great extent in light of the fact that the robotized annotations (model forecasts) are regularly not absolutely right, the mislabeled samples (i.e., error or clamor) are possibly considered when refreshing training models and successively collected in the subsequent learning iterations, prompting a progressive lessening of model execution [3]–[1][2]. The event of this issue should profoundly identify with two elements: model goodness, and rightness and decent variety of selected data when refreshing training data [3]. An ineffectively performed model diminishes the unwavering quality of the computerized annotations, and builds the danger of including mislabeled samples into the refreshed training set. Moreover, with regards to the characteristic expectation tendency of a model, the decent variety of selected data in SSL may be constrained [2], [4]. Including progressively selected data from one model presumably prompts a higher befuddled appropriation between the refreshed training set and test set [2].

To address the presentation debasement issue of SSL, numerous endeavors have been made with regards to machine learning. In [20] and [3], Cohen et al. utilized unlabeled data to scan for a superior structure of Bayesian Network. This calculation can adequately lighten the issue, yet it is just intended for probabilistic models. In [35], Nigam et al. recommended to allocate various loads to unlabeled data as per their forecast probabilities (i.e., certainty). Their methodology at that point prepares another model utilizing the blend of unique labeled and new weighted-unlabeled data, and emphasizes. This strategy adequately lessens the inconvenient impact of ineffectively labeled data by machines [5]. Further, as opposed to such a delicate weighted strategy, its parallel rendition was much of the time utilized too. That is, just a couple of most unhesitatingly anticipated data are added to the labeled data set [3]. Moreover, another enhanced form was presented by Li et al. [6], by which the unlabeled data are effectively related to the assistance of some nearby data in a local diagram. By doing this, it shields those mislabeled data from being added to the training set; henceforth, a less uproarious training set is acquired [6].

In this article, we propose a novel SSL approach called enhanced collaborative SSL (ecSSL), with the reason to address the exhibition corruption issue by leveraging multiple modalities and models with a re-assessment process on selected data. Contrasted and past work, the proposed approach would utmost be able to redesign the integrity of the recognition model just as the 'rightness' and assorted variety of selected data. When all is said in done, our principle commitments can be summed up as follows.

- We abuse the corresponding of multiple modalities (i.e., sound and video) and characterization models for SSL. This mix is critical and accepted to offer at any rate two advantages: to manufacture an enhanced and vigorous emotion recognition model, and to choose increasingly precise and various data in the SSL procedure. Exploiting multiple models is initially roused by the work introduced in where diverse machine learning models can be adapted commonly.

- We propose to consecutively reconsider recently selected data to build the rightness of selected data. It should address conceivably mislabeled data in past iterative learning stages and this further improves the general certainty of the framework expectations.

- We show the predominance of the proposed ecSSL approach on a multimodal database and give clever examination.

II. RELATED WORK

Emotion recognition is the way toward recognizing human emotion. Individuals change generally in their precision at perceiving the emotions of others. Utilization of innovation to help individuals with emotion recognition is a generally early exploration zone. For the most part, the innovation works best on the off chance that it utilizes multiple modalities in setting. Until this point in time, the most work has been directed on mechanizing the recognition of outward appearances from video, verbally expressed articulations from sound, composed articulations from text, and physiology as estimated by wearables.

Humans show a lot of inconstancy in their capacities to perceive emotion. A key point to remember when learning about robotized emotion recognition is that there are a few wellsprings of "ground truth," or truth about what the genuine emotion is. Assume we are attempting to perceive the emotions of Alex. One source is "the thing that would the vast majority state that Alex is feeling?" For this situation, 'reality' may not relate to what Alex feels, yet may compare to what a great many people would state it resembles Alex feels. For instance, Alex may really feel dismal, yet he puts on a major grin and afterward the vast majority state he looks glad. On the off chance that a robotized technique accomplishes indistinguishable outcomes from a gathering of spectators it might be viewed as precise, regardless of whether it doesn't really quantify what Alex genuinely feels. Another wellspring of 'truth' is to ask Alex what he genuinely feels. This works if Alex has a decent feeling of his inner state, and needs to mention to you what it is, and is equipped for articulating it precisely or a number. Be that as it may, a few people are alexithymic and don't have a decent feeling of their interior sentiments, or they can't impart them precisely with words and numbers. When all is said in done, getting to reality of what emotion is really present can take some work, can fluctuate contingent upon the measures that are selected, and will for the most part include keeping up some degree of vulnerability.

Semi-supervised learning is a way to deal with machine learning that joins a limited quantity of labeled data with a lot of unlabeled data during training. Semi-supervised learning falls between unsupervised learning (with no labeled training data) and supervised learning (with just labeled training data).

Unlabeled data, when utilized related to a modest quantity of labeled data, can create impressive improvement in learning exactness. The procurement of labeled data for a learning issue frequently requires a talented human operator (for example to interpret a sound section) or a physical trial (for example deciding the 3D structure of a protein or deciding if there is oil at a specific area). The expense related with the naming procedure in this way may render enormous, completely labeled training sets infeasible, while obtaining of unlabeled data is generally reasonable. In such circumstances, semi-supervised learning can be of incredible useful worth. Semi-supervised learning is likewise of hypothetical enthusiasm for machine learning and as a model for human learning.

A lot of freely indistinguishably appropriated models with comparing marks and unlabeled models are prepared. Semi-supervised learning consolidates this data to outperform the arrangement execution that can be acquired either by disposing of the unlabeled data and doing supervised learning or by disposing of the marks and doing unsupervised learning.

Semi-supervised learning may allude to either transductive learning or inductive learning.[1] The objective of transductive learning is to derive the right marks for the given unlabeled data as it were. The objective of inductive learning is to derive the right mapping from to .

Naturally, the learning issue can be viewed as a test and labeled data as test issues that the educator understands for the class as a guide in taking care of another arrangement of issues. In the transductive setting, these unsolved issues go about as test questions. In the inductive setting, they become practice issues of the sort that will make up the test.

It is superfluous (and, as per Vapnik's standard, hasty) to perform transductive learning by method of deriving an order rule over the whole information space; in any case, practically speaking, calculations officially intended for transduction or acceptance are frequently utilized conversely.

III. PROPOSAL WORK

Enhanced Collaborative Semi-Supervised Learning

One primary disadvantage of SSL is error accumulation, as referenced in Section I. For conventional SSL, the data selected by the machine are completely trusted and pooled into the training set. Be that as it may, a portion of these data are unavoidably mislabelled by and by, and bring about a loud training set (cf. Area I). To handle this issue, we propose to not generally trust the automatically labeled data, and call this methodology enhanced SSL. The pseudocode portraying the calculation is appeared in Algorithm 1. The center thought of this methodology is to hold the recently selected data in the first unlabelled data set at each learning iteration. In doing this, the recently selected data will be reconsidered by the accompanying enhanced model. In this way, it is conceivable to address mislabelled data in future iterations with an improved model. Normally, the recently selected samples may not be selected again in the accompanying learning process, I. e., S I 6⊂ Sj , I < j. In particular, given the gradual number of selected samples per learning iteration n, the I-th learning iteration will choose I × n samples altogether, while the unlabelled data assortment U remains the size of nu, for our situation.

Initialise:
n_l: number of initial labeled training samples;
n_u: number of unlabeled samples;
n: incremental number of selected samples per learning iteration;
h: classification model;
$\mathbf{x}$: feature set, i. e., $\mathbf{x}_a$, $\mathbf{x}_v$, or $\mathbf{x}_{av}$

for $i = 1, \ldots, I$ **do** *% iterate learning process*
  Train classifier $h^i := f(\mathcal{L}^i(\mathbf{x}, y))$;
  Predict $(y'_{\mathbf{x}}, E(y'_{\mathbf{x}})) \leftarrow h^i(\forall \mathbf{x} \in \mathcal{U})$;
  % re-evaluate the whole original unlabeled set
  Split $\mathcal{U} = \{\mathcal{U}^c, c = 1, \ldots, C\}$, where $\forall \mathbf{x} \in \mathcal{U}^c$, $y'_{\mathbf{x}} = c$;
  for $c = 1, \ldots, C$ **do** *% equally selected per class by the strategy of minimum entropy*
    Set $n^i = i \times \lfloor n/C \rfloor$;
    Copy $\mathcal{S}^c$ from $\mathcal{U}^c$, $size(\mathcal{S}^c) = n^i$, and satisfy $E(y'_{\mathbf{x}^c}) \leq E(y'_{\mathbf{x}'^c})$; $\forall \mathbf{x}^c \in \mathcal{S}^c \quad \forall \mathbf{x}'^c \in (\mathcal{U}^c \setminus \mathcal{S}^c)$
    $\mathcal{S}^i = \bigcup \mathcal{S}^c$;
  end
  $\mathcal{L}^{i+1} = \mathcal{L}^0 \cup \mathcal{S}^i$;
end

Algorithm 1
Enhanced Semi-Supervised Learning

Methodology Based Collaborative SSL

The proposed collaborative SSL (cSSL) in this article can be viewed as an expansion of co-training, where the perspectives include not just the component spaces (i.e., methodology based cSSL), yet in addition the recognition models (i.e., model-based cSSL, examined in Section II-D). At the point when coordinated with the enhanced SSL, the new calculation is named as enhanced cSSL.

The pseudocode portraying the calculation of enhanced cSSL dependent on multimodality is shown in Algorithm 2. Contrasted and self-training, methodology based cSSL (e.g., sound, video, text, and physiology) utilizes multiple modalities as autonomous 'sees' for training various models. Contrasted and co-training, it can actualize multiple, instead of two, modalities in the learning framework, which is like multi-see learning with less limitation as far as restrictive autonomy (For additional subtleties, the peruser is alluded to [2]).

Initialise:
n_l: number of initial labeled training samples;
n_u: number of unlabeled samples;
n: incremental number of selected samples per learning iteration;
h: classification model

for $i = 1, \ldots, I$ **do** *% iterate learning process*
  • *either based on* ***multi-modality***
  for $p = 1, \ldots, P$ **do** *% use P* modalities
    Train classifier based on the p-th modality, $h^{ip} := f(\mathcal{L}^i(\mathbf{x}_p, y))$;
    Classification $(y'_{\mathbf{x}_p}, E(y'_{\mathbf{x}_p})) \leftarrow h^{ip}(\forall \mathbf{x}_p \in \mathcal{U})$;
  end
  Merge predictions $y'_{\mathbf{x}} \leftarrow M(y'_{\mathbf{x}_1}, \cdots, y'_{\mathbf{x}_p})$;
  Average entropies $\bar{E}(y'_{\mathbf{x}}) \leftarrow \frac{1}{P}\sum_{p=1}^{P} E(y'_{\mathbf{x}_p})$;

  • *or based on* ***multi-model***
  for $q = 1, \ldots, Q$ **do** *% use Q* models
    Train the q-th classifier $h^{iq} := f_q(\mathcal{L}^i(\mathbf{x}, y))$;
    Classification $(y'^q_{\mathbf{x}}, E(y'^q_{\mathbf{x}})) \leftarrow h^{iq}(\forall \mathbf{x} \in \mathcal{U})$;
  end
  Merge predictions $y'_{\mathbf{x}} \leftarrow M(y'^1_{\mathbf{x}}, \cdots, y'^Q_{\mathbf{x}})$;
  Average entropies $\bar{E}(y'_{\mathbf{x}}) \leftarrow \frac{1}{Q}\sum_{q=1}^{Q} E(y'^q_{\mathbf{x}})$;

  Split $\mathcal{U} = \{\mathcal{U}^c, c = 1, \ldots, C\}$, where $\forall \mathbf{x} \in \mathcal{U}^c$, $y'_{\mathbf{x}} = c$;
  for $c = 1, \ldots, C$ **do**
    Set $n^i = i \times \lfloor n/C \rfloor$;
    Copy $\mathcal{S}^c$ from $\mathcal{U}^c$, $size(\mathcal{S}^c) = n^i$, and satisfy $\bar{E}(y'_{\mathbf{x}^c}) \leq \bar{E}(y'_{\mathbf{x}'^c})$; $\forall \mathbf{x}^c \in \mathcal{S}^c \quad \forall \mathbf{x}'^c \in (\mathcal{U}^c \setminus \mathcal{S}^c)$
    $\mathcal{S}^i = \bigcup \mathcal{S}^c$;
  end
  $\mathcal{L}^{i+1} = \mathcal{L}^0 \cup \mathcal{S}^i$;
end

Algorithm 2: Enhanced Collaborative Semi-Supervised Learning based on Multi-Modality or Multi-Model.

Furthermore, as opposed to traditional co-training where various perspectives independently select the samples that are arranged with most minimal entropies and afterward combine them (i.e., least individual-entropy strategy) [15], [41], cSSL takes a base joint-entropy strategy. That is, all forecasts acquired by different perspectives for each example will be converged as one by methods for dominant part casting a ballot. Especially, in the even cases, an official conclusion is allocated to the classification grouped with the least entropy. This calculation improvement can not just maintain a strategic distance from the expectation strife brought about by various perspectives yet in addition conceivably increment the robotized annotation rightness of the selected data [43]. Besides, the last entropy is determined by averaging all entropies got by various perspectives. These consolidated

expectations and entropies will be then depended on for the accompanying data choosing activity.

CONCLUSION

To use the universal unlabeled data for automatic emotion recognition, this article proposed enhanced collaborative Semi-Supervised Learning (SSL). Not at all like conventional SSL, it plays out a data re-assessment process on recently selected data (enhanced strategy) on one hand. Then again it takes a common learning process among multiple modalities and models (collaborative strategy). The proposed approaches have been efficiently assessed on the generally utilized audiovisual full of feeling database RECOLA in different settings. The trial results show that the proposed approaches essentially improve the framework execution by upgrading the accuracy and decent variety of selected data. All the more as of late, profound learning calculations have pulled in colossal consideration and made an extraordinary progress with regards to machine learning. This will frame one of the primary examination headings later on, by considering various profound learning structures in the SSL systems.

REFERENCES

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 32– 80, Jan. 2001.

[2] M. Chen, P. Zhou, and G. Fortino, "Emotion communication system," IEEE Access, vol. 5, pp. 326–337, Dec. 2016.

[3] M. S. Hossain and G. Muhammad, "An emotion recognition system for mobile applications," IEEE Access, vol. 5, pp. 2281–2287, Feb. 2017.

[4] B. G. Lee, T. W. Chong, B. L. Lee, H. J. Park, Y. N. Kim, and B. Kim, "Wearable mobile-based emotional response-monitoring system for drivers," IEEE Transactions on Human-Machine Systems, vol. 47, no. 5, pp. 636–649, Oct. 2017.

[5] Y. Kim, T. Soyata, and R. F. Behnagh, "Towards emotionally-aware AI smart classroom: Current issues and directions for engineering and education," IEEE Access, vol. PP, no. 99, pp. 1–1, Jan. 2018.

[6] A. Halevy, P. Norvig, and F. Pereira, "The unreasonable effectiveness of data," IEEE Intelligent Systems, vol. 24, no. 2, pp. 8–12, Mar. 2009.

[7] J. Deng, S. Frhholz, Z. Zhang, and B. Schuller, "Recognizing emotions from whispered speech based on acoustic feature transfer learning," IEEE Access, vol. 5, pp. 5235–5246, Mar. 2017.

[8] Z. Zhang, N. Cummins, and B. Schuller, "Advanced data exploitation for speech analysis: An overview," IEEE Signal Processing Magazine, vol. 34, no. 4, pp. 107–129, July 2017.

[9] S. Hantke, T. Appel, F. Eyben, and B. Schuller, "iHEARu-PLAY: Introducing a game for crowdsourced data collection for affective computing," in Proc. International Conference on Affective Computing and Intelligent Interaction (ACII), Xi'an, China, 2015, pp. 891–897.

[10] B. Wu, E. Zhong, D. H. Hu, A. Horner, and Q. Yang, "SMART: Semisupervised music emotion recognition with social tagging," in Proc. SIAM International Conference on Data Mining (SDM), Austin, TX, 2013, pp. 279–287.

[11] M. Schels, M. Kachele, M. Glodek, D. Hrabal, S. Walter, and ¨ F. Schwenker, "Using unlabeled data to improve classification of emotional states in human computer interaction," Journal on Multimodal User Interfaces, vol. 8, no. 1, pp. 5–16, Mar. 2014.

[12] M. Giulianelli, "Semi-supervised emotion lexicon expansion with label propagation and specialized word embeddings," arXiv preprint arXiv:1708.03910, Aug. 2017.

[13] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," Speech Communication, vol. 53, no. 9, pp. 1062–1087, Dec. 2011.

[14] Z. Zhang, F. Weninger, M. Wollmer, and B. Schuller, "Unsupervised ¨ learning in cross-corpus acoustic emotion recognition," in Proc. IEEE workshop on Automatic Speech Recognition and Understanding (ASRU), Big Island, HI, 2011, pp. 523–528.

[15] Z. Zhang, J. Deng, and B. Schuller, "Co-training succeeds in computational paralinguistics," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada, 2013, pp. 8505–8509.

[16] Z. Zhang, E. Coutinho, J. Deng, and B. Schuller, "Cooperative learning and its application to emotion recognition from speech," IEEE Transactions on Audio, Speech, Language Processing, vol. 23, no. 1, pp. 115–126, Jan. 2015.

[17] S. Li, C.-R. Huang, G. Zhou, and S. Y. M. Lee, "Employing personal/impersonal views in supervised and semi-supervised sentiment classification," in Proc. Annual Meeting of the Association for Computational Linguistics (ACL), Uppsala, Sweden, 2010, pp. 414–423.

[18] J. Deng, X. Xu, Z. Zhang, S. Frhholz, and B. Schuller, "Semisupervised autoencoders for speech emotion recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 1, pp. 31–43, Jan. 2018.

[19] A. Mahdhaoui and M. Chetouani, "Emotional speech classification based on multi-view characterization," in Proc. IEEE International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, 2010, pp. 4488–4491.

[20] I. Cohen, N. Sebe, F. G. Cozman, M. C. Cirelo, and T. S. Huang, "Learning bayesian network classifiers for facial expression recognition both labeled and unlabeled data," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Madison, WI, 2003, p. 595.

[21] I. Cohen, N. Sebe, F. G. Cozman, and T. S. Huang, "Semi-supervised learning for facial expression recognition," in Proc. ACM SIGMM international workshop on Multimedia Information Retrieval (MIR), New York, NY, 2003, pp. 17–22.

[22] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[23] M. Nicolaou, H. Gunes, and M. Pantic, "Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space," IEEE Transactions on Affective Computing, vol. 2, no. 2, pp. 92–105, Apr. 2011.

[24] S. Petridis and M. Pantic, "Prediction-based audiovisual fusion for classification of non-linguistic vocalisations," IEEE Transactions on Affective Computing, vol. 7, no. 1, pp. 45–58, Jan. 2016.

[25] O. Celiktutan and H. Gunes, "Automatic prediction of impressions in time and across varying context: Personality, attractiveness and likeability," IEEE Transactions on Affective Computing, vol. 8, no. 1, pp. 29–42, Jan. 2017.

[26] J. Han, Z. Zhang, F. Ringeval, and B. Schuller, "Prediction-based learning for continuous emotion recognition in speech," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2017, pp. 5005–5009.

[27] J. Han, Z. Zhang, M. Schmitt, M. Pantic, and B. Schuller, "From hard to soft: Towards more human-like emotion recognition by modelling the perception uncertainty," in Proc. ACM on Multimedia Conference (MM), Mountain View, CA, 2017, pp. 890–897.

[28] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," IEEE Transactions on Affective Computing, vol. 3, no. 2, pp. 211–223, Apr. 2012.

[29] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres Torres, S. Scherer, G. Stratou, R. Cowie, and M. Pantic, "AVEC 2016: Depression, mood, and emotion recognition workshop and challenge," in Proc. International Workshop on Audio/Visual Emotion Challenge (AVEC), Amsterdam, Netherlands, 2016, pp. 3–10.

[30] X. Zhu, "Semi-supervised learning literature survey," Department of Computer Sciences, University of Wisconsin at Madison, Madison, WI, Tech. Rep. TR 1530, 2006.

[31] D. Yarowsky, "Unsupervised word sense disambiguation rivaling supervised methods," in Proc. Annual meeting of the Association for Computational Linguistics (ACL), Stroudsburg, PA, 1995, pp. 189–196.