2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops): Work in Progress

OCEAN: Towards Developing an Opportunistic Continuous Emotion Annotation Framework

Akhilesh Adithya*, Snigdha Tiwari*, Sougata Sen*, Sandip Chakraborty†, Surjya Ghosh*
*Department of Computer Science and Information Systems, BITS Pilani Goa, INDIA
†Department of Computer Science and Engineering, IIT Kharagpur, INDIA
Email: {f20190044, f20190555, sougatas, surjyag}@goa.bits-pilani.ac.in, sandipc@cse.iitkgp.ac.in

2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) | 978-1-6654-1647-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/PerComWorkshops53856.2022.9767359

***Abstract*—Emotion-aware video consumption services typically deploy a machine learning model to infer the emotion automatically and provide the service accordingly. The ground truth labels to train such models are usually collected as emotion *self-reports* from users in a continuous manner (using additional devices) while they watch different videos. This process of continuous annotation induces additional cognitive workload and degrades the viewing experiences. To overcome these challenges, we propose a framework *OCEAN* (Opportunistic Continuous Emotion Annotation) that collects emotion self-reports opportunistically. The key idea of OCEAN is to identify the moments when the physiological responses change significantly and use *only* those moments as the self-report collection (or probing) moments. We evaluate OCEAN using the CASE dataset (a publicly available dataset capturing continuous emotion annotation for different videos). Our preliminary results demonstrate that OCEAN reduces continuous annotation effort (median number of the probe is four and an average reduction of 89% probes) and yet collects ratings similar to continuous annotations.**

***Index Terms*—Emotion self-report, Opportunistic annotation**

I. Introduction

Recently, there has been a surge in the use of emotion-aware video recommendation and consumption services [1] that triggered various exciting applications, including improved learning experience in MOOC platforms [2], quantifying emotional reactions to advertisements [3], and enhancing interaction quality with videos [4], etc. Such services typically use machine learning (ML)-based models at the backend to infer users' emotions and adapt the content of the video accordingly. A large amount of ground truth emotion labels are required to pre-train such ML models; therefore, the typical approach is to collect the ground truth data through human-based annotations in the form of *self-reports* during the video consumption. However, human-based emotion label annotation has two fundamental limitations. First, frequent probing for annotations disrupt the viewing experience of the users [5], and second, because of such degraded viewing experience, the labels can be noisy and erroneous. Therefore, practical researches and service developments in this direction need an efficient method for emotion annotations. Interestingly, emotions are subjective; therefore, we cannot avoid human-in-the-loop altogether for emotion annotations.

Currently, different annotation strategies are used to collect emotion self-reports. The most widely adopted approach is the post-interaction or post-stimuli one, where the participants provide emotion self-reports based on a standard scale (e.g., Self-assessment Manikin (SAM) [6]) after watching the video. However, these approaches fail to capture intra-video subtle nuances present in the videos. For example, a video can embed different emotions (e.g., *happiness*, *anger*, *sadness*), but in the post-stimuli approach, capturing and time-aligning all the emotions is challenging. To address these issues, researchers use continuous emotion annotation strategies, where participants provide emotion annotations using a mouse or joystick or another similar device, as they watch the video [7]. For example, in the CASE (*Continuously Annotated Signals of Emotion*) dataset [8], participants continuously provided emotion annotation (*valence* and *arousal* based on the Circumplex Model of emotion [9]) using a joystick. Works such as FEELTRACE [7] have also collected continuous emotion annotations using such devices. For such continuous emotion annotation, the user needs to focus on two jobs simultaneously – (i) watching the video and (ii) annotating emotion labels using the joystick or similar devices, thus introducing a significant amount of cognitive workload. Indeed, such overheads not only affect the viewing experience but also impact the emotion labels. This paper explores whether an opportunistic annotation strategy that collects self-reports only at the relevant moments (when the emotional reactions change) can alleviate the need for continuous annotation and mitigate the associated cognitive overloads.

Two different factors guide the key idea of developing such an opportunistic annotation strategy. First, human emotions usually persist for a period once felt; this is known as the persistent effect of emotion [10]. Second, in a video, the emotional nature of content does not change very frequently (e.g., in every continuous frame) [11]. Therefore, it may not be essential to collect the annotations continuously. Rather an annotation strategy that automatically decides the opportune moment of emotion variation (based on the variation in the physiological responses) and requests (or probes) user for self-reports *only* at these points can reduce the cognitive workload and improve the viewing experience. Accordingly, we, in this paper, propose the OCEAN (***O**pportunistic **C**ontinuous **E**motion **AN**notation*) framework to opportunistically identify the probing moments instead of continuous emotion annotation. Using a change-point detection algorithm, the framework observes the physiological signals from different modalities and detects the probing moments based on the abrupt changes in the signal values. To optimize the probing moments fur-

ther, OCEAN clusters the change-point scores (using k-means clustering) and selects points with significant changes in the physiological signals. This approach not only helps to reduce the annotation effort but also collects annotations at those points when the emotion variation has occurred.

We evaluate OCEAN using the publicly available continuous emotion annotation dataset, CASE [8]. The dataset consists of physiological responses and continuous emotion ratings from 30 participants watching different videos. We segment the physiological responses into small windows and use these as input to the framework to identify if the current window is suitable for self-report collection (or probing) based on the changes in the signal values (from the previous window). We observed that for every video, the median number of self-reports to be responded to by the users is less than (or equal to) 4. On average, users need to respond to 89% fewer probes in the framework. We also demonstrate that despite fewer probes, the sampled annotations are very similar to the continuous ones (closely follow the minimum, maximum, and median obtained from the continuous scores), thus indicating the possibility of reducing annotation overhead by probing opportunistically without compromising on annotation quality.

II. Dataset

The CASE dataset [8] collects emotion annotations based on the Circumplex model of emotion [9] in two dimensions (valence, arousal) using a Joystick based Emotion Reading Interface (JERI). At the same time, it tracks the body's physiological reactions in response to the video being shown. The dataset captures physiological responses for different emotional stimuli (amusement, boredom, relaxation, scary). The participants viewed two videos in each of these emotion categories and were asked to provide their emotional responses continuously (in real-time, as they watch the videos) using a Joystick-based interface. The participants moved the Joystick on a 2D plane to record valence and arousal scores (on a scale of 1 to 9) based on the position of the Joystick. The annotation data were sampled at 20Hz. The videos were selected to ensure that all the four quadrants of the emotion Circumplex model were well represented. We show the details of these videos in Table I. As the participants watched the videos, the following physiological signals were continuously collected – Electrocardiograph (ECG), Blood Volume Pulse (BVP), Galvanic Skin Response (GSR), Respiration (RSP), Skin Temperature (SKT), and Electromyography (EMG). These physiological signals were synchronized and were sampled at 1000Hz.

Video id	Emotion	Valence	Arousal	Duration (in sec.)
1	amusing	med/high	med/high	185
2	amusing	med/high	med/high	173
3	boring	low	low	119
4	boring	low	low	160
5	relaxing	med/high	low	145
6	relaxing	med/high	low	147
7	scary	low	high	197
8	scary	low	high	144

TABLE I: Details of the videos present in the CASE dataset for continuous emotion annotation collection.

Thirty volunteers (15F, 15M) aged between 22 and 37 years watched these videos and recorded their emotional and physiological responses following a within-subject study design. To avoid carry-over effects, the order of the videos in a session were modified in a pseudo-random manner, such that the resulting video sequences varied between participants. To isolate the emotional response elicited by the different videos, they were interleaved by a two-minute long blue screen. This two-minute period also allowed the participants to rest in-between annotating the videos.

III. OCEAN Framework

In this section, we discuss the OCEAN framework as shown in Fig. 1. The key idea of the proposed framework is to identify the points where the physiological responses changed during the video consumption and to probe users only at these points. Accordingly, we divide the framework into following key stages – (a) *physiological response segmentation*, (b) *probing moment detection* (based on change score), and (c) *probing moment optimization*. Next, we discuss each of the steps.

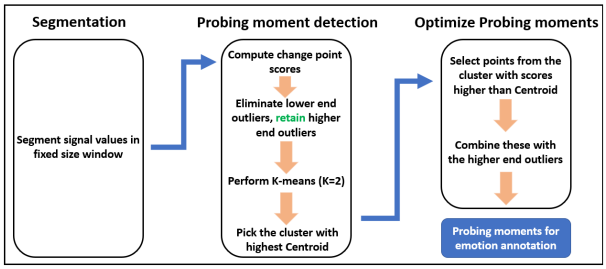


Fig. 1: OCEAN Framework for identifying opportune probing moments based on physiological response variation

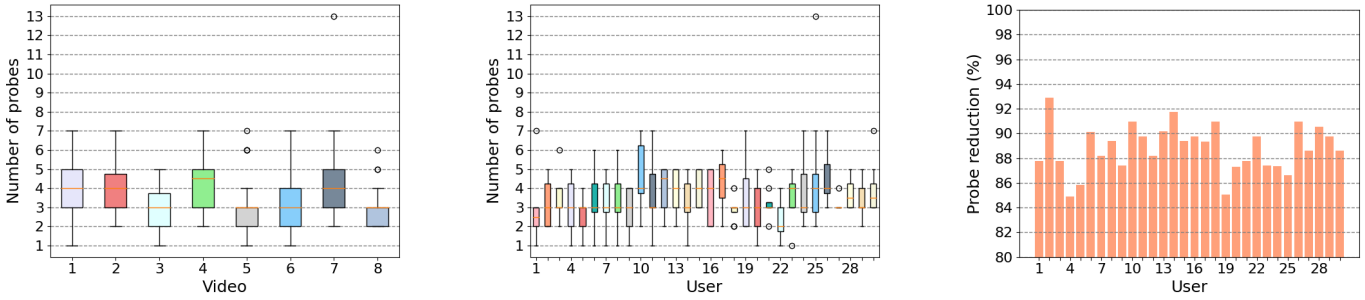
A. Physiological Signal Segmentation

In this stage, we segment the physiological responses into fixed size windows. Although the dataset contains physiological responses from different signals (BVP, ECG, EMG, GSR, respiration, and skin temperature), we did not consider ECG signal in this analysis as ECG signal is relatively noisy [12]. For every participant and video combination, we segment the remaining physiological responses into fixed size windows. We segment the physiological signals into 5-seconds windows (derived empirically) and use it as input to the next module.

B. Probing Moment Detection

We apply the change point detection algorithm [13] on the windows to detect the opportune probing moments. These algorithms compute a score (known as change point score) to identify any abrupt changes in the time-series values (in this case physiological responses) of two consecutive data segments (or windows). As a result, the variation in the physiological responses manifested by the emotional changes will be reflected by a high score.

The probing moment detection steps are as follows. First, we compute the change point scores for every two consecutive windows of signal values captured from a video using the RuLSIF algorithm [14]. Second, we apply an outlier elimination



(a) Video-wise probe (b) User-wise probe (c) User-wise probe reduction

Fig. 2: Number of probes using OCEAN framework - (a) video-wise probes, (b) user-wise probes, (c) user-wise probe reduction using OCEAN with respect to continuous self-report collection.

strategy on the set of change point scores. Specifically, we eliminate outliers (change point score) at the lower extreme (lesser than 3*standard deviation (σ) than mean (μ)) as they indicate minimal change (or no change) in the physiological signal values and thus are not ideal for probing. However, the outliers at the higher end (greater than $\mu + 3 * \sigma$) are *retained separately* as these points indicate very high change (and therefore suitable for collecting emotion self-reports). Third, all the *remaining points* (change point scores) are grouped using the k-means clustering algorithm [15]. The value of k is set to 2 so that the scores indicating a change are in one cluster, and no changes are grouped into another cluster. Intuitively, the change point scores are higher for the points (between two consecutive windows) where the physiological signals have actually changed. Therefore, we pick the cluster whose centroid has a higher value as the cluster of interest for probing. The scores in this cluster are the set of candidates for probing users for the self-report. These steps are performed for every user and every video, as individual responses may vary based on the content of different videos.

C. Probing Moment Optimization

We aim to optimize the number of probing moments because probing a user in all candidate moments may lead to interruption. To achieve this, we apply the following filtering strategy. First, from the cluster of higher centroid, we select only those points having a score higher than the centroid (this reduces the number of probing moments). Next, We combine these points with the retained higher-end outliers (as discussed in Section III-B) to obtain the final set of change points, where we can probe the user for self-report collection.

IV. Evaluation

This section evaluates OCEAN in terms of annotation effort reduction, quality of sampled annotation, and ability in capturing the emotion response variations for timely probing.

A. Annotation Effort Reduction

We evaluate the efficiency of reducing continuous annotations using the OCEAN framework. To validate this, we identify the number of probes issued for each of the videos in Fig. 2a. We observe that we need to probe four times on average for each video. Similarly, we check the number of probes for every user in Fig. 2b. In this case, we observe that for most of the users ($\approx 93\%$), the median number of probes that OCEAN requires is four across all videos. We also find the reduction in probing rate using OCEAN. To identify this, for every user, we compute the total number of probes (for all videos) that would have been triggered if the annotations were collected at every five seconds (as OCEAN also employs a five seconds window) interval – n_{act}. Similarly, we compute the total probes issued for the same user using the OCEAN framework for all the videos – n_{ocean}. Thereafter, the probing reduction is computed as $\frac{(n_{act}-n_{ocean})\times 100}{n_{act}}$. We present the reduction in probing rate for all the users in Fig. 2c. For each user, we obtain at least 85% reduction in probing rate and, on average, a reduction of 89% using OCEAN. These findings demonstrate that continuous annotations can be avoided, and the self-reports can be collected opportunistically by probing users a few times only. We next discuss whether this reduction in probing rate impacts the annotation quality.

B. Emotion Annotation Quality

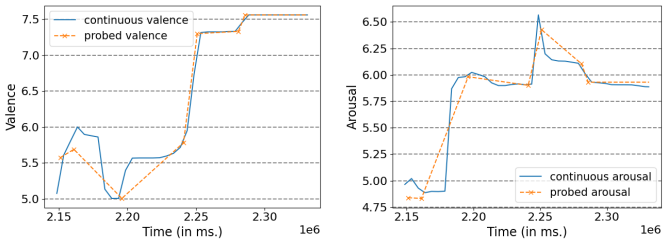
To compare the quality of sampled annotation with continuous annotations, we compare several statistical measures (maximum, minimum, and median) of sampled values (valence and arousal) and continuous values (valence and arousal). We present these statistical measures of valence and arousal scores for every video based on sampled annotations and continuous annotations in Table II. We observe that for all the videos, the statistical measures for both sampled and continuous annotations are almost identical (except video 6 and 8 maximum valence values). These findings highlight that the sampled annotations are similar to the continuously collected emotion annotations and the framework does not influence the quality of opportunistically collected labels for majority of the cases.

C. Detecting Emotion Response Changes

To ensure that the sampled annotations (for valence and arousal) follow the continuous ratings provided by the users, OCEAN should probe when user's emotion has actually changed. To verify this pattern, we qualitatively compared the continuous annotations and sampled annotations (the ratings

Video id	Valence						Arousal					
	Min (Act.)	Min (Sam.)	Max (Act.)	Max (Sam.)	Med (Act.)	Med (Sam.)	Min (Act.)	Min (Sam.)	Max (Act.)	Max (Sam.)	Med (Act.)	Med (Sam.)
1	1.763	1.883	9.5	9.5	6.266	6.284	3.164	3.552	9.013	9.013	5.248	5.624
2	3.058	3.514	9.5	9.5	6.169	6.562	1.71	2.936	8.414	8.396	5.008	5.101
3	2.42	2.967	7.39	7.39	5	5	0.509	0.509	7.005	6.375	3.772	4.844
4	2.686	3.145	7.943	7.572	5.009	5.05	0.5	1.402	6.477	6.477	3.836	4.065
5	1.163	1.8	9.5	9.5	6.052	5.845	0.5	0.5	6.291	5.937	4.115	4.337
6	2.717	2.981	9.206	7.705	5.397	5.702	2.905	2.912	6.83	6.609	5	5
7	0.5	0.5	7.813	7.513	2.887	3.2	1.023	1.023	9.5	9.5	7.200	7.711
8	0.5	0.5	8.919	6.871	4.613	3.198	3.018	4.287	9.5	9.5	6.45	7.417

TABLE II: Comparison of valence and arousal scores for all the videos based on continuous annotation (denoted as Act.) and the sample annotation (denoted by Sam.) using OCEAN framework. Min., Max., and Med. denote minimum, maximum, and median values respectively. For all the videos, all these values (min, max, and med) are similar for continuous and sampled annotation for valence and arousal.



(a) Continuous vs probed valence (b) Continuous vs probed arousal

Fig. 3: Comparing continuous and probed annotation for one representative user (user 4) and one video (video 1) - (a) valence comparison, (b) arousal comparison.

as collected based on the probing moments) for every user and every video combination. It reveals that for most cases, probed annotation moments closely follow the continuous annotation (see Fig. 3 for one representative user's valence and arousal for a video). However, there are a few users (e.g., user 14, 30), for whom this pattern is not very identical. We envision that individual physiological response variation could have attributed to this, something we plan to investigate in future.

V. Conclusion and Future Work

This paper presents OCEAN, a framework for opportunistically collecting emotion self-reports for videos rather than continuous annotation. OCEAN detects significant variations in the physiological signals (as the participants watch videos) using a change-point detection algorithm and identifies a set of relevant self-report collection moments using the k-means algorithm. The preliminary findings over the publicly available CASE dataset highlight that OCEAN reduces the continuous annotation overhead, yet records annotations similar to the continuous ratings. However, while developing OCEAN, we came across some potential opportunities and challenges that can be explored in the future versions of this work.

Latency of Probing: One of the critical factors in the performance of OCEAN is the window size used for change-point detection. For an online system, this will indicate the minimum window for which the system needs to collect the sensing data to decide whether there is any change. Albeit, it is implausible that a subject's emotional state changes across such a small window. A thorough investigation needs to be done to quantify the errors introduced in the system.

Storage Reduction: One of the essential advantages that OCEAN provides, in addition to the reduction of probes, is the overall reduction of storage space required to log the continuous sensed data. Notably, OCEAN achieves this by precisely observing the changes in the physiological sensors, which potentially indicate the opportune moments where the emotional state may have changed.

VI. Acknowledgements

We sincerely thank Soumyajit Chatterjee for his valuable inputs and suggestions, which helped to improve the paper. The works of Dr. Sandip Chakraborty is supported through IMPRINT-I Research Grant (Project ID: 6720) via Sanction order F. No.: 41-2/2015-T.S.-I (Pt.),Dt.09-01-2016.

References

[1] Y. Deldjoo, M. Schedl, P. Cremonesi, and G. Pasi, "Recommender systems leveraging multimedia content," *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–38, 2020.

[2] Y. Zhao, T. Robal, C. Lofi, and C. Hauff, "Stationary vs. non-stationary mobile learning in moocs," in *Adjunct publication of the 26th UMAP*, 2018, pp. 299–303.

[3] P. Pham and J. Wang, "Attentivevideo: A multimodal approach to quantify emotional responses to mobile advertisements," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 9, no. 2-3, pp. 1–30, 2019.

[4] J. McNally and B. Harrington, "How millennials and teens consume mobile video," in *Proceedings of the 2017 ACM TVX*, 2017, pp. 31–39.

[5] T. Zhang, A. El Ali, C. Wang, A. Hanjalic, and P. Cesar, "Rcea: Real-time, continuous emotion annotation for collecting precise mobile video ground truth labels," in *ACM CHI*, 2020, pp. 1–15.

[6] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[7] R. Cowie, E. Douglas-Cowie, S. Savvidou*, E. McMahon, M. Sawey, and M. Schröder, "'FEELTRACE': An instrument for recording perceived emotion in real time," in *ITRW Speech-Emotion*, 2000.

[8] K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, and F. Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," *Scientific data*, vol. 6, no. 1.

[9] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[10] P. Verduyn and S. Lavrijsen, "Which emotions last longest and why: The role of event importance and rumination," *Motivation and Emotion*, vol. 39, no. 1, pp. 119–127, 2015.

[11] S. Wang and Q. Ji, "Video affective content analysis: a survey of state-of-the-art methods," *IEEE Transactions on Affective Computing*, vol. 6, no. 4, pp. 410–430, 2015.

[12] S. L. Joshi, R. A. Vatti, and R. V. Tornekar, "A survey on ecg signal denoising techniques," in *2013 International Conference on Communication Systems and Network Technologies*. IEEE, 2013, pp. 60–64.

[13] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowledge and information systems*, vol. 51, no. 2, pp. 339–367, 2017.

[14] S. Liu, M. Yamada, N. Collier, and M. Sugiyama, "Change-point detection in time-series data by relative density-ratio estimation," *Neural Networks*, vol. 43, pp. 72–83, 2013.

[15] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the royal statistical society. series c (applied statistics)*, vol. 28, no. 1, pp. 100–108, 1979.