Where's Your Mind At? Video-Based Mind Wandering Detection During Film Viewing

Angela Stewart, Nigel Bosch, Huili Chen, Patrick J. Donnelly, & Sidney K. D'Mello
University of Notre Dame
384 Fitzpatrick Hall, Notre Dame, IN, 46556, USA
{astewa12, pbosch1, hchen6, pdonnel4, sdmello}@nd.edu

ABSTRACT

Mind wandering (MW) is a ubiquitous phenomenon in which attention involuntarily shifts from task-related processing to task-unrelated thoughts. This study reports preliminary results of a video-based MW detector during film viewing. We collected training data in a study where participants self-reported when they caught themselves MW over the course of watching a 32.5 minute commercial film. We trained classification models on automatically extracted facial features and bodily movement and were able to detect MW with an F_1 of .30. The model was successful in reproducing the MW distribution obtained from the self-reports.

Keywords

Mind wandering, facial features, user modeling, affective computing, film viewing

1. INTRODUCTION

Most of us have had the experience of engaging in an activity, such as such as reading or watching a film, only to suddenly realize that our attention has gradually drifted away from task-related thoughts to completely unrelated thoughts like dinner or weekend plans. This shift in attention is known as mind wandering (MW). Considerable research over the last decade has documented MW's widespread incidence during a host of real-world activities. For example, in one large-scale study MW was tracked in 5,000 individuals from 83 countries working in 86 occupations with an iPhone app that prompted people to report their thoughts at random intervals throughout the day [4]. People reported MW for 46.9% of the prompts, which confirmed numerous lab studies on the pervasiveness of MW, which is estimated to occur approximately 20-50% of the time, depending on the person, task, and the environmental context [4, 5].

In addition to being quite frequent, MW is also detrimental to performance across a number of tasks, such as reading comprehension, signal detection, memory recall, and retention of learned content [7]. Further, the negative correlation between MW and performance increases in proportion to task complexity [7]. When compounded with its high frequency, MW can have serious consequences on performance and productivity. Therefore, we believe that next-generation intelligent interfaces could benefit from some mechanism to detect and address MW. Of course, an interface must first detect MW before it can respond to it. Thus, the goal is to develop a fully-automated video-based detector of MW during film viewing.

2. METHOD

We used data from an existing study [5] in which 107 participants viewed the narrative film "The Red Balloon" (1956, Figure 1) while a video of their faces and upper bodies was recorded with a commercial webcams. Participants self-reported MW by pressing keys when they caught themselves "thinking about anything else besides the movie" or "thinking about the task itself but not the actual content of the movie."

Figure 1. Screenshot of "The Red Balloon"

MW self-reports were sparsely distributed throughout the 32.5 minute video. Our first task was to create data instances corresponding to short windows of time preceding MW reports. The procedure for creating instances was as follows:

1) Add a 3-second offset before the self-caught MW report to account for movement due to reporting (i.e., the key press).
2) For all MW reports that are within *S* seconds of each other, where *S* is the segment size, only keep the first MW report and remove any others.
3) Partition the video between consecutive MW reports into $(t_i - t_{i-1}) / S$ segments, where t_{i-1} and t_i are the timestamps of consecutive MW reports. The segment immediately preceding the MW report at t_i is a MW segment. All other segments between t_{i-1} and t_i are not MW segments.
4) Extract features from a window of data of size *w*, where $w < S$, from the end of each segment generated in step 3.
5) The remaining time $(S - w)$ seconds in the segment is the gap that is not analyzed.

The procedure described above is depicted in Figure 2 using a 45-second window size. In this study, we chose a 55 second segment length as it resulted in a MW rate of approximately 20% to 25%, which was consistent with previous research [1, 4, 5]. We explored various windows sizes within the 55-second segment and chose a 45-second window size for this initial analysis. We generated a total of 2,734 segments, after excluding instances in which the participants' faces could not be registered in the frame to the extent that less than 1 second of data could be extracted from the 45-second window.

UMAP '16, July 13-17, 2016, Halifax, NS, Canada
ACM 978-1-4503-4370-1/16/07.
http://dx.doi.org/10.1145/2930238.2930266

We used FACET [9], a commercialized version of the CERT computer vision software, for facial feature extraction. FACET provides likelihood estimates of the presence of 19 action units as well as head pose, face position, and face size. Features were created by aggregating FACET estimates in the window using maximum, median, and standard deviation for aggregation. In all, there were 75 facial features which were complemented by 3 features that measured gross body movement from the videos [8].

Several standard classifiers from Weka [3] were used to discriminate between MW and not MW instances. We applied SMOTE (on training data only) to account for data imbalance [2]. Feature selection was performed on a subset of participants in the training set. We evaluated the performance of our classifiers using leave-one-participant-out cross-validation.

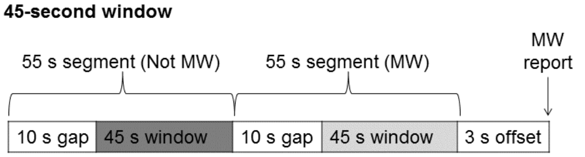

Figure 2. Example of window segmentation approach, using 45-second window sizes. Features are extracted from the dark grey (Not MW) and light grey (MW) windows.

3. RESULTS

The most accurate model was a support vector machine (SVM) classifier [6]. We compared the distribution of per participant MW rates as predicted by the model (Figure 3 - middle) to the distribution of self-reported MW rates (Figure 3 - top). We note that the model was quite accurate at predicting when participants had zero or low MW rates (compare points A and C in Figure 3) but over predicted MW in a large number of participants (compare points B and D in Figure 3). This resulted in an average predicted MW rate of double the self-reported rate.

To address this, we adjusted the model's threshold of when to predict MW. Originally, any instance that exceeded a confidence of .500 was classified as MW. We adjusted this threshold to .600, which yielded the distribution shown at the bottom of Figure 3. The resultant model no longer over predicted MW (i.e., compare points B and F in Figure 3), and correctly predicted when participants had zero or low MW rates (points A and E in in Figure 3). This model had a MW prediction precision of .30, recall of .30, and consequently a F_1 score of .30 for the MW class (minority class).

4. CONCLUSION

This present study demonstrated the feasibility of using facial features to detect MW during film viewing. Our approach used a setup that required affordable and accessible equipment to detect MW in an everyday context. We were moderately successful in advancing a fully-automated system for automatic MW detection with evidence for generalizability to new users. The ubiquity of webcams have opened up the possibility of advancing research in attentional state estimation, thereby enabling an entirely new generation of attention-aware interfaces.

5. ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation (NSF) (DRL 1235958 and IIS 1523091). Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF.

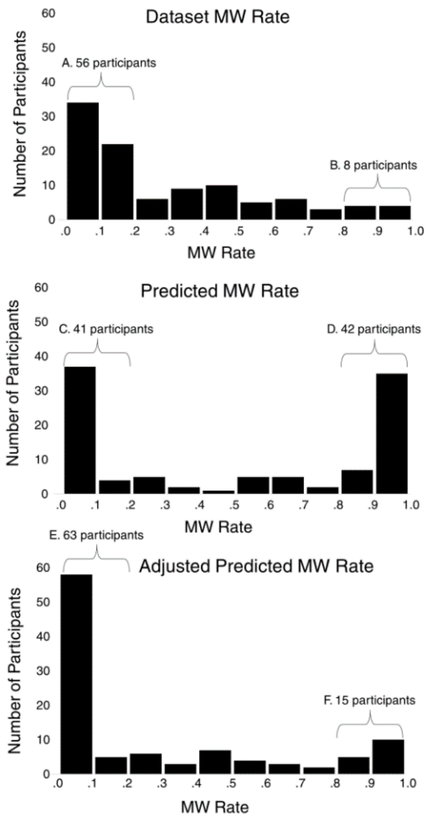

Figure 3. MW rate distributions The self-reported MW rates of the dataset (top), predicted MW rates (middle), and adjusted predicted MW rates (bottom) are shown.

6. REFERENCES

[1] Bixler, R. and D'Mello, S. 2014. Toward fully automated person-independent detection of mind wandering. *Proceedings of the 22nd International Conference on User Modeling, Adaptation, and Personalization* (Switzerland, 2014), 37–48.

[2] Chawla, N.V. et al. 2002. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. (2002), 321–357.

[3] Holmes, G. et al. 1994. Weka: A machine learning workbench. *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on* (1994), 357–361.

[4] Killingsworth, M.A. and Gilbert, D.T. 2010. A wandering mind is an unhappy mind. *Science*. 330, 6006 (2010), 932–932.

[5] Kopp, K. et al. 2015. Mind wandering during film comprehension: The role of prior knowledge and situational interest. *Psychonomic bulletin & review*. (2015), 1–7.

[6] Platt, J. 1998. Fast training of support vector machines using sequential minimal optimization. *Advances in Kernel Methods - Support Vector Learning*. MIT Press. 41 – 64.

[7] Randall, J.G. et al. 2014. Mind-Wandering, cognition, and performance: A theory-driven meta-analysis of attention regulation. *Psychological bulletin*. 140, 6 (2014), 1411.

[8] Westlund, J.K. et al. 2015. Motion Tracker: Camera-Based monitoring of bodily movements using motion silhouettes. *PloS one*. 10, 6 (2015).

[9] 2016. *Emotient module: Facial expression emotion analysis*.