

"The smell of fear" - In Search of Patterns Between Emotions and Breath Composition - Data Cleaning

Jason Tam¹, Musa Abdel-Rahman², Catherine Liu³
Supervised by Jörg Wicker

I. ALIGNMENT AND EXCLUSION OF SESSIONS

As the objective of this study is to determine if potential associations are present between the scene labels and the gas content detected in theatre, data collected from different sessions are grouped by the movie played for analysis.

With the relationship between CO₂ and human breath composition being the best understood compared to other available gas channels of the detector, data from the CO₂ channel is used to determine the general effect from the audience. Figure 1 shows the data from the CO₂ channel collected from all sessions of the movie 'Buddy'. It is clear that while the traces for most of the sessions followed a general pattern, a small number of them has distinctive difference in shape. Some of the differences are emphasized in the normalized data shown in Figure 2.

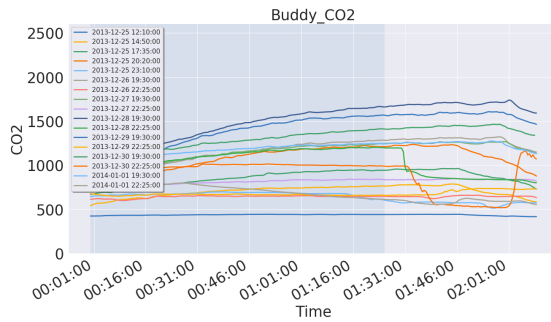


Fig. 1. Uncleaned sessions data of 'Buddy'

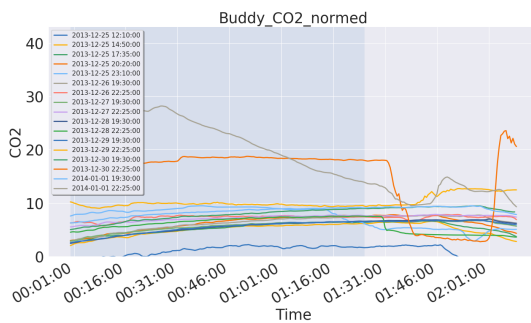


Fig. 2. Uncleaned and normalized sessions data of 'Buddy'

The starting time for all of these sessions plotted in the graph are aligned with the 'begin' time provided in the data. The shaded area represent the duration of the movie using the aligned starting time, with the ending time estimated by the number of columns that represent 30 second intervals in the scene label data table. Data of extended time periods (45 minutes) after the estimated ending time are included to more accurately determine the ending time of the movie, as it can be safely assumed that delays that affect the true starting time of the movie are not uncommon. It is also assume that a noticeable drop in CO₂ counts would occur as the audience leave the theatre after the end of the movie, and this is used to align the playing time of the movie across different sessions. The aligned data for the 'Buddy' movie is shown in Figures 3 and 4.

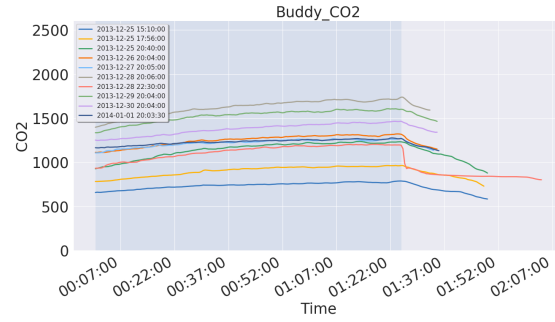


Fig. 3. Cleaned sessions data of 'Buddy'

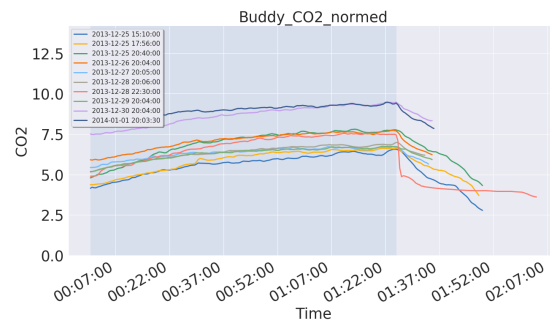


Fig. 4. Cleaned and normalized sessions data of 'Buddy'

¹jtam013@aucklanduni.ac.nz

²mabd292@aucklanduni.ac.nz

³cliul29@aucklanduni.ac.nz

This data cleaning strategy seems to be appropriate as there are still a good number of remaining sessions for further analysis. Similar results were obtained for 'The

Hunger Games: Cathcing Fire' and 'Walter Mitty', as shown respectively in Figures 5 and 6.

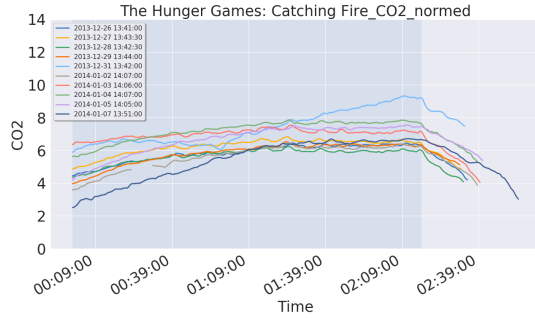


Fig. 5. Cleaned sessions data of 'The Hunger Games: Cathcing Fire'

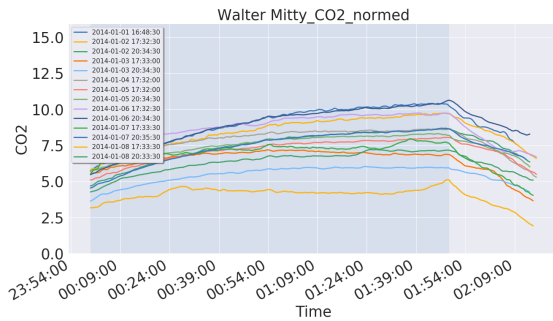


Fig. 6. Cleaned and normalized sessions data of 'Walter Mitty'

However the same cannot be said with the other movies. 'Machete Kills' only has two sessions, and only one of the two sessions can be aligned with the same logic, as the CO₂ count in the other trace starts to noticeably drop before reaching the length of the movie. The two sessions are shown in Figure 7, and whether the aligned session has enough value to keep for the analysis is open for discussion.

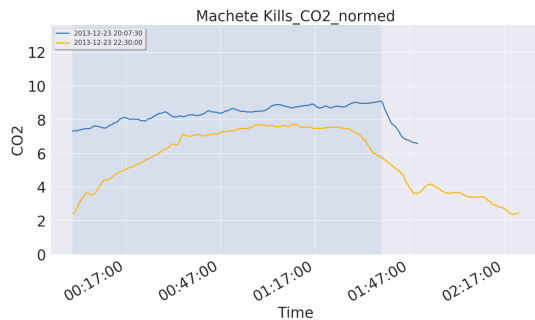


Fig. 7. Cleaned and normalized sessions data of 'Machete Kills'

Even with 45 minutes of extended time considered, noticeable drop in counts for the CO₂ channel does not appear in any sessions of 'Hobbit 2' and 'Paranormal Activity'. These are shown in Figure 8 - 11.

Particularly for 'Paranormal Activity', there seem to be two main types of patterns for different sessions, possibly due

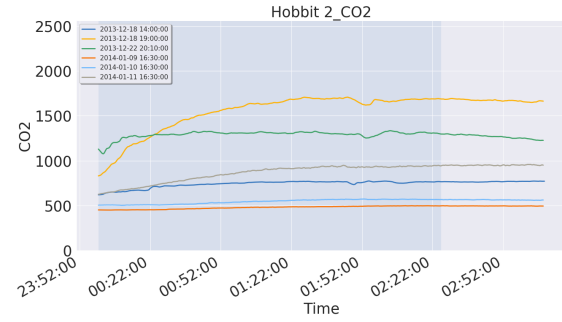


Fig. 8. Uncleaned sessions data of 'Hobbit 2'

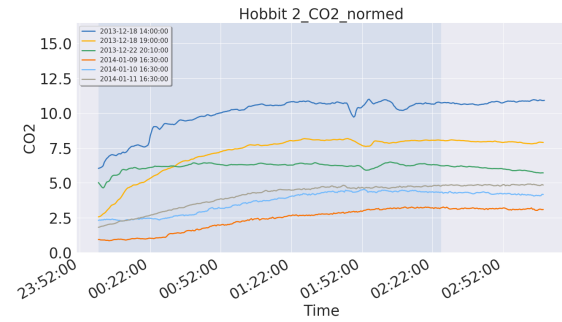


Fig. 9. Uncleaned and normalized sessions data of 'Hobbit 2'

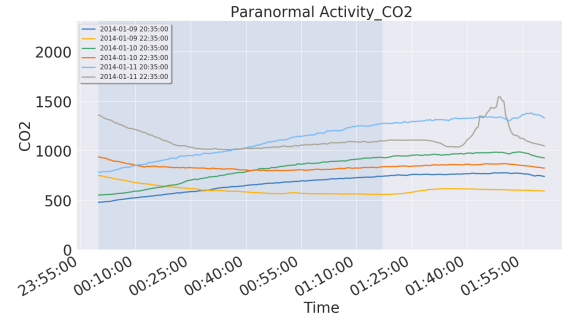


Fig. 10. Uncleaned sessions data of 'Paranormal Activity'

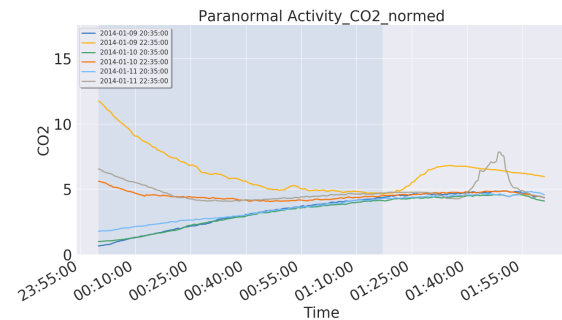


Fig. 11. Uncleaned and normalized sessions data of 'Paranormal Activity'

to the aftermath of the sessions before. Whether the sessions of these two movies have enough value to keep for the analysis is also open for discussion, while I can look further into the extended time periods to look for a similar drop in counts to hopefully perform the same alignment. I have excluded all three of these movies ['Machete Kills', 'Hobbit 2', 'Paranormal Activity'] for now and produced an updated CSV file for clustering analysis. The dataset contained within should be the cleanest subset available, whether the quantity is adequate for the analysis can be examined from the results obtained, and the strategy can be further refined if needed.