Quantifying Changes in Student Behaviors During COVID-19 with Non-negative Matrix Factorization

Keywords: human behavior, behavioral change, diversity, routine modelling, nmf

Extended Abstract

In this work, we analyze human behavior by identifying qualitative differences in students' daily routines. We use time diaries collected from students' smartphones in two different time periods (2018 and 2020), to understand how behavioral routines changed during the pandemic. Using their reported activities we extract meaningful daily routines with *Non-negative Matrix Factorization* (NMF) to understand typical behaviors in 2018 or 2020. Then, we construct a behavioral space with the NMF extracted components to quantify the differences between these typical behaviors.

Dataset Collection The data we use combines datasets from two studies [1, 2], collected from an Italian university for a period of two weeks in 2018 and two weeks in 2020. The 2020 data collection happened during the second wave of the covid pandemic, but with relaxed restrictions: students of this study had to attend lessons mostly in presence. In both studies, students downloaded a smartphone application, which stored sensor data and allowed them to fill *Time Diaries*. The application prompted a questionnaire every 30 minutes to the students, who had to select and report the activity they were doing, their location, whether they had company and their mood.

Methods We organized the data so that for each student the daily Time Diaries follow circadian cycles starting at 5AM. Then, we organized the users data from both datasets in 3 different matrices, one for each type of information that was collected (location, social, activity). The matrices contain the normalized counts of the observed behavior in a specific timeslot for each student. To extract meaningful daily routines from these matrices, we applied Non-negative Matrix Factorization (NMF), which decomposes the behavior matrix X so that $X \approx WK$, with K corresponding to the extracted components, weighting each of the features, and W representing the weights of each component per student. We chose the number of components both considering the *cophenetic correlation coefficient*, which measures how the distance between different samples is preserved in the NMF transformation, and the interpretability of the components. Moreover, we identified the most important components to distinguish behavior from 2018 and 2020, with *Information Gain*. Finally, we used these components to create a behavioral space in which we could measure the distance between students' behaviors.

Results An initial exploratory analysis of the datasets of students in 2018 and 2020 showed how the activities in the two groups were similar. We looked at the complexity of the average student behavior by computing their entropy on the behavior matrices, and there was no difference between 2018 and 2020 for activities. When looking at the other information collected in the Time Diaries, instead, both location and sociality had lower entropy in 2020, indicating that students shared less time with other people and their visit patterns became more predictable in that year. In fact, we observe that home locations percentage increased significantly from 66.7% in 2018 to 86.3% in 2020.

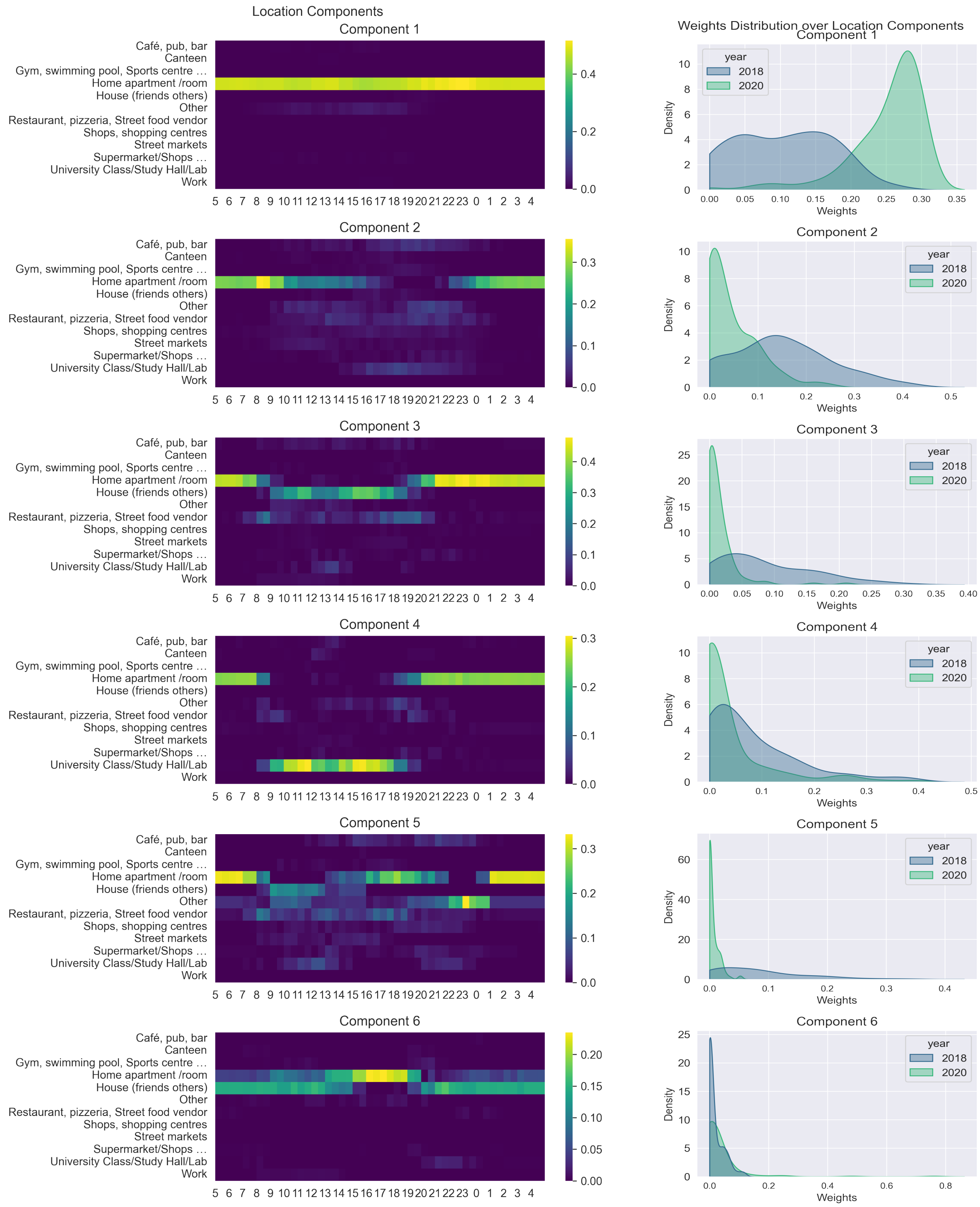
To look at changes in behavior more in depth, we applied NMF to the activities, location and sociality information. Figure 1a shows the components extracted for the locations, which represent common students' routines. For example, the first component corresponds to a day spent at home, while component four represents a day spent at the university. Figure 1b, instead, shows the distribution of the weights associated with the NMF components for each student. This information helps us to understand which are the components and the behaviours that changed the most between the two groups. In particular, components 4 and 6, which represent days spent at the university and days spent at home (either own home or friends' homes) respectively, have similar distributions for 2018 and 2020, meaning that this kind of behavior did not change drastically from one year to the other. Components 1 and 5, which represent a day at home (more common in 2020), and a day spent in other locations such as friends' homes or restaurants (more common in 2018), have different distributions meaning that students in 2020 stayed more at home and visited less places. We quantified the importance of the components by computing their *information gain*, measuring how informative component weights are in distinguishing samples from the 2018 or 2020 datasets. The results are coherent with the shapes of the weights distributions and the most informative components for the locations are 1, 3 and 5. This pattern of spending more time at home is also present in the sociality aspect of the data, for which the extracted components showed how being alone has high weights for 2020 students' behaviors and being with friends and/or classmates has very low weights for 2020 behaviors and higher weights for 2018 behaviors. This can also be explained by the covid restrictions that were in place at the time of 2020 data collection: students could not enjoy much time outside in the city, but other activities such as bars and restaurants were open. The student sample had lessons in presence, but study spaces were closed. Given these restrictions, the shift of the social behavior seems natural. Finally, the activities were not much informative in distinguishing the two types of behavior. In fact, only the first component, which showed students moving, having social life and doing other activities, has a significant information gain. This is due to the fact that there is less variability in the daily behavior pattern in 2020. This lack of difference supports the hypothesis that, despite the restrictions, students did not change much their activity habits, but they changed how the sociality part was present, spending more alone time and in private homes rather than in public places.

To further support these results, we quantitatively measure the distance between students' behaviors by embedding them in a multimodal *behavioral space*, spanned by the most informative components. In this behavioural space, two separate clusters emerge, which separate 2018 and 2020 behaviors.

References

[1] Fausto Giunchiglia, Enrico Bignotti, and Mattia Zeni. Personal context modelling and annotation. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 117–122, 2017.

[2] Fausto Giunchiglia, Ivano Bison, and Matteo Busso et al. A worldwide diversity pilot on daily routines and social practices (2020-2021). *University of Trento Technical Report*, (#DISI-2001-DS-01), 2021.



(a) NMF Location components (b) Weights distribution

Figure 1: Results obtained by applying our method to the locations data. In Figure 1a, we can see the 6 NMF extracted components, with the x axis representing the time of day (divided in 30 minutes bins) and the y axis representing the different locations. The heatmap shows the weight of a location in a specified timeslot and show different daily routines. In Figure 1b, we can see the distribution of the weights for the two datasets and how much the individual components contribute to the original behavior. For example, since the distribution of 2020 weights has a distribution including higher values for the first component, in which students stay at home, we can say that component 1 is more important for reconstructing 2020 behavior