Mikko Hyyryläinen

Type: *Taking a paper of your choice which is not using data science methods and introduce how you would use data science approaches to redo that paper 10 points per paper*

Link to the article: https://journals.sagepub.com/doi/10.1177/1749975518791431

The article that I am "redoing" is Marie Plessz's (2019) *Is Cooking Still a Part of Our Eating Practices? Analysing the Decline of a Practice with Time-Use Surveys* which was published in the journal *Cultural Sociology* (13:1). The original article utilizes linear regression, which is sometimes classified as a machine learning method, but it seems that it is quite often classified as belonging to "classic statistics" which is considered to be more or less from the practices of machine learning. Therefore, the question of how this article could be redone with machine learning methods is justified.

The goal of the article was to "examine the extent to which changes in the association between cooking durations and the number of eating events at home account for the decline in household cooking time" (94). The topic is framed in the terms of practice theory and following Shove et al. (2012) the authors define cultural change as consisting of changes in a set of practices (94). The data they use comes from time-use surveys where people produce 24-hour diaries of their activities and the time spent with them (97) and they consider these time-use surveys as "temporal footprints" of practices (98). They clarify their explanatory task as "we wanted to explore how trends in food preparation duration can be accounted for by changes in the association between cooking durations and eatings in time-use surveys" (98). In addition, the study was a comparison between the US and France. As a summary we can say that the general observation that the article is based on is that the time that is used in cooking has declined and the authors are interested in explaining why this is the case. The possible explanations are 1) the decline is caused by the decrease in eating events at home (that is eating practices) and 2) the practices of cooking (for example, faster food preparation etc) have changed.

The rationale for "redoing" this paper is the way that the authors approach culture and practices. Their conception of culture implies that they consider culture as something that is unified national culture where all the members of that culture share the same practices. There is nothing wrong with this conception and it is quite conventional conception. However, culture can also be considered as something that is somewhat fragmented and consists of various sub-cultures, or, various sets of practices that differ from each other. The idea that I have for redoing this paper is based on this fragmented view of culture and the machine learning method that seems to be able to catch this fragmentation is K-Means clustering.

The question I would ask is *how the changes in the structure of the culture of food/cooking are related to "the decline in household cooking time"*. The rationale for this question comes

from the idea that from a culture it is possible to find various different cooking practices – i.e. we do not only consider that there are "national cooking practices". However, we do not probably know in advance how many different sub-cultures we can find from national data. Hence, clustering seems to be a fitting machine learning method for this task. The logic then goes that, since the variables we use are evidence about culture and its practices, clusters should then be evidence about different sub-cultures. We could then ask if changes in the structure of culture is somehow related to the decline in household cooking time.

The data that was used in the article consisted of four sets of data. Two French time-use surveys: from 1988 and 2009–2010. Two American time-use surveys: from 1985 and 2010. Having four sets of data makes it possible to do four different cluster analyses and also do comparisons between the countries and cultures. The first two clusterings would be done with the data sets from the 1980's. The goal of the cluster analysis would be to use the elbow-method and "human evaluation" in find the right number of clusters that would answer the question of "from what kind of subcultures these national cultures consist of". For example, in what kind of sub-cultures food preparation takes a lot of time and do this kind of sub-cultures differ from each other in some other way. We could also get the answer of what kind of sub-culture, or cluster, has the most cases in it.

The second clustering would be about the data sets from the 2000's. In this phase we could do two things. First task would be to do clustering analysis with the number of clusters that was considered to describe the data from the 1980's in the best way. This could tell us two interesting things about the culture. Firstly, if the number that was used would produce clusters that "make sense" also in the data from the 2000's, then we would know that the structure of the culture had not changed. However, could perhaps be a change in how many cases are in each cluster. Secondly, if it would not produce clusters that "make sense", then we would know that the structure of culture has changed in some way. Finding a higher number of clusters that would "make sense" would indicate that the structure of the culture would have become more fragmented. If the number would be lower, then we could say that the culture has become more unified. We would also be able to compare the trajectories of national cultures.

How this then would be different from the study reported in the article (apart from the method that is used)? Article's discussion section starts with the statement "French and American households devote less time to domestic food preparation than they once did. Is cooking still part of their eating practices?". This indicates that what they are interested about is what is happening in some kind of average household and therefore linear regression is a good tool for this and the answer for the proposed question is based on the relationships that variables have with each other. My idea of redoing this paper with clustering attempts to map if the way culture and its structure has (probably) changed could be considered to be the reason for the decline in the time that is used in domestic food preparation. It is not necessarily a better way to explain the decline but it does account for the fragmented structure

of a culture. In addition, it could produce knowledge on how cultures change (or do not change) over time. However, I am not sure if the data that is used would be a good foundation for K-means clustering.