6 Identifying physical activity profiles in COPD patients using topic models

With the growing amount of physical activity (PA) measures, the need for methods and algorithms that automatically analyse and interpret unannotated data increases. In this work, PA is seen as a combination of multimodal constructs that can co-occur in different ways and proportions during the day. The design of a methodology able to integrate and analyse them is discussed, and its operation is illustrated by applying it to a dataset comprising data from COPD patients and healthy subjects acquired in daily life. The method encompasses different stages. The first stage is a completely automated method of labelling low-level multimodal PA measures. The information contained in the PA labels are further structured using topic modelling techniques, a machine learning method from the text processing community. The topic modelling discovers the main themes that pervade a large set of data. In our case, topic models discover PA routines that are active in the assessed days of the subjects under study. Applying the designed algorithm to our data provides new learnings and insights. As expected, the algorithm discovers that PA routines for COPD patients and healthy subjects are substantially different regarding their composition and moments in time in which transitions occur. Furthermore, it shows consistent trends relating to disease severity as measured by standard clinical practice.