Challenges for the Similarity-Based Comparison of Human Physical Activities Using Time Series Data

Tomasz Szczepanski[1], Kerstin Bach[1] and Agnar Aamodt[1]

Department of Computer and Information Science
Norwegian University of Science and Technology, Trondheim, Norway
http://www.idi.ntnu.no

Abstract. In this position paper we present various aspects of comparing human physical activities using time series data within the SELF-BACK project. The goal of the project is to develop a decision support system for patients suffering from non-specific low back pain. The system will give users advice in form of a self-management plan that is based on self-reported physical and psychological symptoms as well as activity stream data collected by a wristband. Here, we discuss the activity stream representation difficulties as well as various challenges that arise when comparing the resulting activity streams.

Keywords: Case-Based Reasoning, Time Series Representation, Data Streams, Similarity Assessment

1 Introduction

SELFBACK [2] is a research project developing a decision support system that is aimed towards helping patients to facilitate, improve and reinforce self-management of non-specific low back pain. The SELFBACK system will constitute a data-driven, predictive decision support system that uses the Case-Based Reasoning (CBR) methodology to capture and reuse patient cases in order to suggest the most suitable activity goals and plans tailored for an individual patient. This will be based on data from two different types of sources. One is a questionnaire presented to the patient at suitable intervals in order to capture self-reported general information and progress of symptoms. The other is a stream of activity data collected using a wristband. The incoming data will be analyzed to classify the patients current state and recent activities, and matched against past cases in order to derive follow-up advices to the patient.

Two of the many challenges during matching and similarity assessment when using the CBR methodology, is developing a suitable abstraction of the wristband activity stream and an adequate comparison algorithm for such streams. The abstractions and the comparison algorithm must catch the important aspects and compare these in a meaningful way relevant in a non-specific low back pain domain.

We will in the next section shortly present the different aspects and challenges that need to be considered in order to develop such a comparison. Specifically, we will focus on the challenges arising when comparing the activity streams of patients.

2   Challenges

The main goal of the SELFBACK project is to give non-specific low back pain patients personalized advice based on their reported symptoms and activity patterns collected by a wristband. Three major challenges arise when comparing the activity streams:

1. The activity stream needs to contain the knowledge needed and used for the treatment of non-specific low back pain.
2. The comparison of such streams must be made in a way that is meaningful when comparing patients with non-specific low back pain.
3. Comparing activity streams with missing data. The wristbands have limited battery time and are often not water proof and needs to be taken off during showering or swimming activities. Users might find it uncomfortable to wear the wristbands continuously and try to loosen the strap or take breaks in wearing them, which might introduce gaps in recordings.

Because these problems are very novel, we have very little knowledge on what to include and how to compare such activity streams. There is some evidence that sleeping patterns and prolonged periods of inactivity are correlated with an increase in non-specific low back pain [1], [3], [5]. There is however, very limited data on what activity types or intensities of activity types are beneficial/harmful when having non-specific low back pain.

The activity streams that we have experimented with can contain activity data from periods of up to several weeks. This is because to see the effect of a change in behaviour, by applying some advice, one often needs to wait for about 4-6 weeks. Including all kinds of information into an activity stream (type of activity, intensity, skin temperature, pulse, moisture of the skin, etc.), comes with an increase in the computational cost when comparing the resulting activity streams against each other. See figure 1 for examples of abstractions that can be included.

We have also experimented with different resolutions on our activity streams. We are able to capture and transform data into 1 second as well as 1 hour windows to name a few. Decreasing the resolution, lowers the computational comparison cost, but important data may be lost during the transformation. Different resolutions can be combined throughout the activity stream to prevent loss of information. For example a bigger window can be used during prolonged periods of inactivity or sleep. On the other hand, during periods of activity where the intensity or activity types changes frequently, a smaller window can be used to capture and conserve that information.
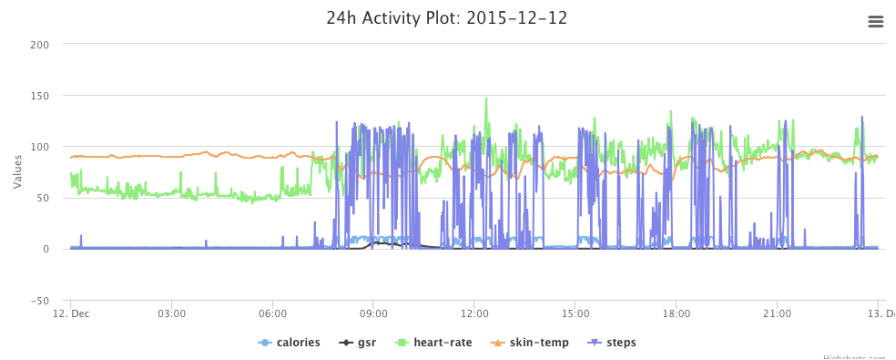
Fig. 1. Three abstractions from collected wristband data: step counts together with heart rate and skin temperature measures.

Other ideas we have been experimenting with include comparing the activity streams based on the activity intensity rather than activity type. Clinicians with experience in the field of non-specific low back pain have suggested, instead of comparing the streams directly, to compare sleeping patterns and periods of prolonged inactivity. This way of comparing activity streams produces very different results than when comparing activity streams containing activity types. This is shown in figure 2.
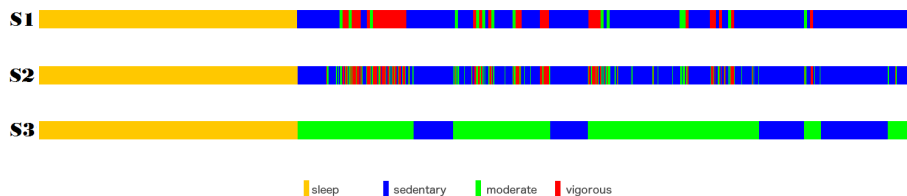

Fig. 2. The data from figure 1 abstracted to a higher level and represented as an activity stream in three different ways. S1 and S2 represents activity intensity throughout the day (S2 with higher resolution). In S3 the activity stream is transformed, as suggested by clinical experts, to only show prolonged inactivity periods. The green periods in S3 represent any activity.

As for the comparison algorithms themselves, we are currently experimenting with a string based approach. We convert the activity stream into a string of characters. When comparing two strings, we compute the percentage of each character per string and the number of consecutive sequences of those characters. We then also compute the longest common sub-sequence, the sequence distance, the number of similar k-mers and the number of unique similar k-mers. The sequence distance is the Levenshtein distance which is the least number of single

character insertions, deletions or substitutions that is required to transform one string into the other [4]. A k-mer is a consecutive substring of length k. All these attributes are combined in to a final similarity metric.

3 Current status and Outlook

We have presented and discussed various challenges that arise when comparing activity streams in the SELFBACK project.

Our current work focuses on the representation and comparison of the activity stream data that is meaningful in the non-specific low back pain domain. We are investigating which time resolution is beneficial and which information should be included in the activity streams. We are also exploring whether and how domain knowledge from non-specific low back pain experts can be incorporated. As shown in figure 2, the representation of the activity streams can differ a lot depending on which information is most relevant (alteration of activities in S1 and S2 vs. the reduction of pro-longed inactivtiy in S3). Therefore we are working on how the representation affects the similarity comparison of activity streams collected from real patients.

Our next steps include the investigation whether activity stream rearrangement could yield towards a more precise comparison. Therefore we investigate if parts of the activity stream are more important than others (for example sleep patterns). When the data collection from the planned randomized control trial is complete, it will also be possible to mine the collected activity streams for behavioral patterns that contribute to an increase or a decrease in non-specific low back pain. For example we could discover that a certain amount of time standing followed by a certain amount of time sitting is increasing non-specific low back pain etc. Finally we will conduct experiments on how much weight the activity stream comparison should have compared to the self-reported reported questionnaire data from the patients.

References

1. Alsaadi, S.M., McAuley, J.H., Hush, J.M., Lo, S., Lin, C.W., Williams, C.M., Maher, C.G.: Poor sleep quality is strongly associated with subsequent pain intensity in patients with acute low back pain. Arthritis Rheumatology 66(5), 1388–94 (2014)
2. Bach, K., Szczepanski, T., Aamodt, A., Gundersen, O.E., Mork, P.J.: Case Representation and Similarity Assessment in the selfBACK Decision Support System. ICCBR 2116 (accepted for publication)
3. Gupta, N., Christiansen, C.S., Hallman, D.M., Korshj, M., Carneiro, I.G., Holtermann, A.: Is Objectively Measured Sitting Time Associated with Low Back Pain? A Cross-Sectional Investigation in the NOMAD study. PLoS ONE 10(3) (2015)

4. Navarro, G.: A guided tour to approximate string matching. ACM Computing Surveys 33(1), 31–88 (2001)
5. van Tulder, M., Becker, A., Bekkering, T., Breen, A., del Real, M.T.G., Hutchinson, A., Koes, B., Laerum, E., Malmivaara, A.: Chapter 3 european guidelines for the management ofacute nonspecific low back painin primary care. European Spine Journal 15(2), s169–s191 (2006)