Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence

Decomposing Activities of Daily Living to Discover Routine Clusters

Onur Yürüten[1] and **Jiyong Zhang**[2] and **Pearl Pu**[1]
[1]Human Computer Interaction Group,[2]Artificial Intelligence Laboratory
École Polytechnique Fédérale de Lausanne (EPFL), Station 14 Lausanne, Switzerland CH - 1015
{onur.yuruten,jiyong,zhang,pearl.pu}@epfl.ch

Abstract

The modern sensor technology helps us collect time series data for activities of daily living (ADLs), which in turn can be used to infer broad patterns, such as common daily routines. Most of the existing approaches either rely on a model trained by a preselected and manually labeled set of activities, or perform micro-pattern analysis with manually selected length and number of micro-patterns. Since real life ADL datasets are massive, such approaches would be too costly to apply. Thus, there is a need to formulate unsupervised methods that can be applied to different time scales. We propose a novel approach to discover clusters of daily activity routines. We use a matrix decomposition method to isolate routines and deviations to obtain two different sets of clusters. We obtain the final memberships via the cross product of these sets. We validate our approach using two real-life ADL datasets and a well-known artificial dataset. Based on average silhouette width scores, our approach can capture strong structures in the underlying data. Furthermore, results show that our approach improves on the accuracy of the baseline algorithms by 12% with a statistical significance ($p < 0.05$) using the Wilcoxon signed-rank comparison test.

Introduction

The recent advances in sensor technology allow us to bring healthcare systems to our everyday lives in the form of pervasive sensors and software. Using these tools, people can quantify their physical activities and internal metabolisms over time (Smarr 2012). Some systems also incorporate simple techniques to deliver correlation information for personal data (Tollmar, Bentley, and Viedma 2012). However, researchers must employ even more sophisticated methods to understand what physical activity patterns people adopt, and whether these patterns cause variations in the level of physical activeness within individuals (intrapersonal differences) or groups of people (interpersonal differences). A pattern analysis on activity routines can help identify such information, and thus enhance the usefulness of pervasive healthcare systems.

The existing methods on ADL analysis either explicitly specify models for a preselected set of activities (Wadley et al. 2008), or analyse and extract features from repetitive micro-patterns (i.e., motifs). The first approach requires expert knowledge, thus it is costly and delivers a restricted understanding of the data. In the second approach, the appropriate granularity for micro-patterns must be exhaustively searched for any given dataset. As such, despite early successes (Bao and Intille 2004; Cook 2010), studies that adopt these approaches report on a limited amount of physical activities, likely monitored in laboratory conditions (Pham, Plötz, and Olivier 2010; Zheng et al. 2013). Thus, there is a need to formulate unsupervised methods that can be applied to different time scales.

We observe that people adopt some activity routines in their daily living, with some possible deviations every day. Based on this observation, we propose a novel approach to analyse time series activity data. We pre-process the time series with a smoothing filter (Hodrick and Prescott 1997) and extract routines and deviations via a sparse and low rank matrix decomposition technique (Lin, Chen, and Ma 2010). We separately cluster the routines and deviations, and then perform a cross product between routine-clusters and deviation-clusters to find the final memberships for each entry.

Our contributions in this paper are as follows:

- Our approach is different from prior work as it is model-free, and it uses the whole time series data as opposed to a subset of motifs or features.
- We propose a novel combination of low rank and sparse matrix decomposition and time warping techniques for activity analysis. To our knowledge, our approach is the first one in the activity analysis studies to incorporate this approach.
- We show, on two real-life datasets of accelerometer data (calorie expenditure and steps) and of different time scales, that our method can capture distinct structures in ADL time series that are associated with different levels of activeness. Furthermore, we show on a well-known synthetic dataset (Keogh and Kasetty 2003) that we can also obtain high accuracy scores on labelled time series data.

Related Work

Activity analysis studies follow two general directions. The first approach constructs a model of some preselected activities, and establishes the fitness of this model through methods such as Bayesian Learning (Zheng and Ni 2012) and Hidden Markov Models (Cook 2010). The obtained models can serve to predict people's house activities (Cook 2010), to group the users based on their activity routines (Zheng and Ni 2012), or to identify common activity routines (Zheng and Ni 2012). Model-based methods are commonly applied on datasets of location and motion sensors. To obtain sound results in their models, researchers study incorporate domain expert knowledge (and perhaps manually annotate the dataset). This requires substantial effort, and constrains the quality of the analysis to the extent of the expert's knowledge ahead of the quality of the dataset.

As an alternative, studies from the second approach extract features from frequently occurring patterns (motifs in other words), and then construct classifiers based on these features. The bioinformatics field spearheads the research on discovering frequent patterns (we refer the readers to the paper of Sandve and Drablos (2006) for an extensive review). Typically each pattern-based activity recognition study proposes a custom motif-detection algorithm (Pham, Plötz, and Olivier 2010; Patel, Hsu, and Lee 2012; Rashidi et al. 2011), while some prefer to directly incorporate state-of-the-art pattern detection algorithms such as random projection (Vahdatpour, Amini, and Sarrafzadeh 2009) and Closet+ (Ali et al. 2008). Subsequently, for classification, studies either apply state-of-the-art supervised learning techniques such as Support Vector Machines, Decision Trees (Patel, Hsu, and Lee 2012) or incorporate custom data structures (like graph-based clustering (Vahdatpour, Amini, and Sarrafzadeh 2009), and routine-tree (Ali et al. 2008)). It is also possible to construct Hidden Markov Models based on the extracted patterns (Rashidi et al. 2011) or apply ensemble learning (Zheng et al. 2013). Motif-based studies obtained empirical success on datasets a large variety of sources: environmental motion sensors, wearable accelerometers, pressure sensors, and medical analysis data (such as blood tests and urinalysis).

Due to the computational complexity of finding motifs, some studies prefer a fixed length and number of motifs (Vahdatpour, Amini, and Sarrafzadeh 2009). Some other studies report that the accuracy (or other quality measures) of the classification and clustering consistently improves as the number of motifs increase (Rashidi et al. 2011). On the other hand, some studies show that clustering the entire set of subsequences does not produce meaningful results (Keogh and Lin 2005). Therefore, the scientists may have to exhaustively search for the optimal length, and the number of motifs in their studies. This, again, may limit the representation capabilities of the systems.

Methods

We summarize the flow of data processing in Figure 1. We pre-process the ADL time series data with a smoothing filter (Hodrick and Prescott 1997) and apply a low rank and sparse decomposition (Candès et al. 2011) to isolate routines (L-Matrix) from the deviations (S-Matrix). We separately cluster L-Matrix and S-Matrix, using Dynamic Time Warping (Keogh and Pazzani 1999) as the distance metric. We use the well-known Silhouette index (Kaufman and Rousseeuw 2009) to determine the optimal number of clusters. We then perform a cross product of the two separate cluster sets to find the final memberships for each day.

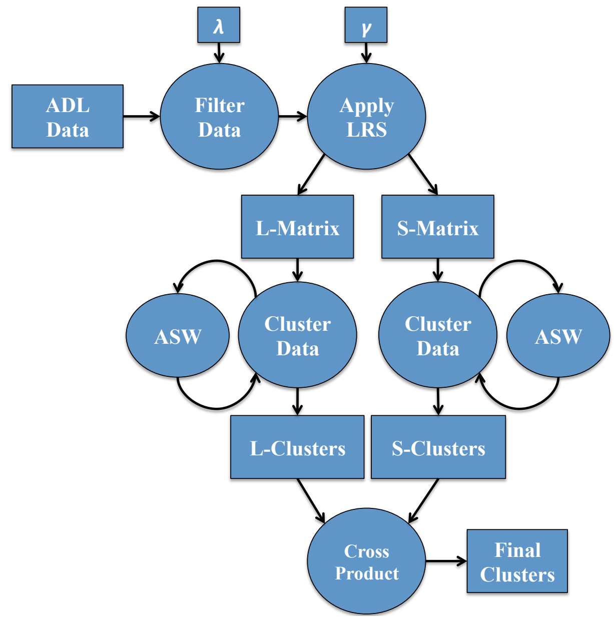

Figure 1: The data flow in our approach. LRS stands for "Low rank and sparse decomposition", and ASW stands for "Average Silhouette Width"

Smoothing Filter

The physical activity time series data may contain noise in the form of small fluctuations. Such characteristics of the raw data can deteriorate the quality of clustering. We address this issue by applying the Hodrick-Prescott filter (Hodrick and Prescott 1997). This is a well-known trend analysis method in economics. The filter decomposes a given time series object $Y = (y_1, ..., y_m)$ into a summation $y_t = T_t + C_t$ such that the objective function

$$\sum_{t=1}^{m} {C_t}^2 + \lambda \sum_{t=2}^{m-1} \left((T_{t+1} - T_t) - (T_t - T_{t-1})\right)^2, \qquad (1)$$

is minimized over $(T_1, ..., T_m)$, where T_t represents the trend component (the desired output), and C_t represents the cyclical component. Increasing the smoothing parameter (λ) results in smoother trend components at a cost of more information loss. We discard the cyclical component and use the trend component in the further steps.

Matrix Decomposition

The *low rank and sparse decomposition* is a recently discovered approach that aims to capture regular and symmetric structures within a possibly corrupted data matrix (Liang et al. 2012). While it is designed for image processing problems such as video surveillance and face recognition, it is also used other high-dimensional data mining tasks such as finding topic models in document analysis (Min et al. 2010).

Based on existing studies (Candès et al. 2011), we can formulate this decomposition problem as

$$min(\|L\|_* + \gamma \|S\|_0) s.t. M = L + S, \quad (2)$$

where L is the low-rank matrix, and S is the sparse matrix. $\|L\|_*$ denotes the nuclear norm of L, which is the best approximation for the rank of L. $\|S\|_0$ is the number of non-zero entries in S. $\gamma>0$ is the parameter to make a trade-off between the rank of L and the sparsity of S. Theoretical studies show that it is optimal to set γ as $1/\sqrt{max(n_1, n_2)}$, where n_1, n_2 are the number of rows and columns of M, respectively (Candès et al. 2011).

The interpretation of the L-matrix and S-matrix differs among the related studies. L is commonly regarded as the "true matrix", which is recovered from the errors and missing values denoted in S (Zhou and Tao 2011). In a related study, L contains linearly aligned images and S contains the rotational errors from the original matrix (Peng et al. 2010). In some other image processing studies, L is considered to be the background and S the non-background objects in the given images (Kyrillidis and Cevher 2012). As such, depending on the application, the information in both of these matrices can be useful.

We use the Linearized Alternating Direction Method (Lin, Chen, and Ma 2010) on the matrix of ADL time series data to identify common daily routines (in the form of low-rank matrix) and deviations (in the form of the sparse matrix). To our knowledge, our study is the first to apply the low rank and sparse decomposition approach to ADL analysis.

Distance Metric: Dynamic Time Warping

ADL routines are subject to nonlinear warps in the time dimensions (e.g., waking up 15 minutes late, having lunch for 30 minutes instead of 45, etc.). Dynamic Time Warping (DTW) is a dynamic programming-based distance metric to compensate these warps (Berndt and Clifford 1994). In contrast to Euclidean distance, DTW takes local misalignments into consideration, and reports the optimal warping path between the given two sequences. The DTW distance between the time series data Q and P can be calculated as

$$DTW(Q, P) = min_W(\sum_{k=1}^{K} d(w_k)), \quad (3)$$

where $d(w_k) = (q_i - p_j)^2$ such that (q_i, p_j) is on the warping path w (Fu 2011). Various studies with artificial datasets (Keogh and Pazzani 1999), image data of letters in historical documents (Rath and Manmatha 2003), speech data (Sakoe and Chiba 1978), and kitchen tool usage data (Pham, Plötz, and Olivier 2010) suggest that DTW improves the classification accuracy of the time series classification algorithms in comparison to Euclidean distance. DTW is sensitive to noise (Fu 2011). This can be overcome by applying additional preprocessing (Rath and Manmatha 2003). We avoid this problem by applying Hodrick-Prescott filter before the matrix decomposition stage.

Clustering

We obtain pairwise distance matrices for L-matrix and the S-matrix. Then we feed these distance matrices to agglomerative hierarchical clustering with complete linkage. As a result, for each row in the original data, there will be one cluster membership from L-matrix and one cluster membership from S-matrix. To determine the final memberships, we perform a cross product of L-clusters and S-clusters, i.e., we explore all possible combinations of L-clusters and S-clusters. The maximum possible number of final clusters is (number of L-clusters) × (number of S-Clusters). We discard the clusters with no members. To guarantee the optimal number of clusters, we select the number of L-clusters and S-clusters that result in the highest average silhouette width.

Experiments

Datasets

CBF Dataset. This artificial dataset (Keogh and Kasetty 2003) contains time series objects that belong to one of three distinct shape characteristics (i.e., Cylinder $c(t)$, Bell $b(t)$ and Funnel $f(t)$, see Figure 2). The dataset can be generated with the following equations:

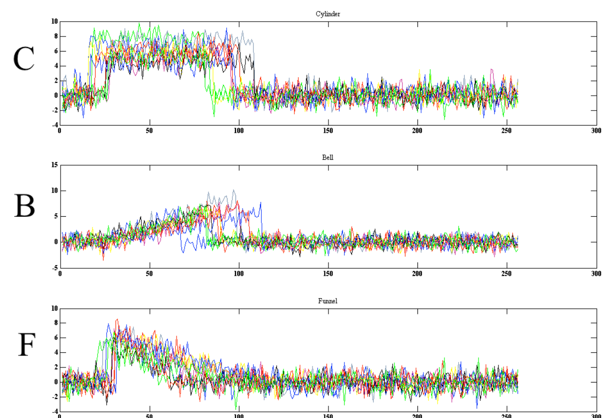

Figure 2: Samples from the CBF dataset. The axes are unitless. Each class of objects (C:Cylinder, B: Bell, F: Funnel) is defined uniquely by its shape characteristics.

$$c(t) = (6 + \eta)\chi_{[a,b]}(t) + \epsilon(t) \quad (4)$$

$$b(t) = (6 + \eta)\chi_{[a,b]}(t)\frac{t - a}{b - a} + \epsilon(t) \quad (5)$$

$$f(t) = (6 + \eta)\chi_{[a,b]}(t)\frac{b - t}{b - a} + \epsilon(t), \quad (6)$$

where η and $\epsilon(t)$ are drawn from a standard normal distribution, a is an integer drawn uniformly from [16, 32], and $b - a$ is an integer drawn uniformly from [32, 96]. We have generated 256 instances for each class (cylinder, bell, and funnel), each of which contains 256 data points.

E-Walk Dataset. This dataset is the courtesy of the Yiqizou company, which provide a platform for people to form social groups and walk together. This dataset contains step counts of 236 people, who wore modern wearable accelerometers in October 2013 for a month. In its raw form, each data point represent activities during a single day. Due to some possible reasons (losing interest in the program, forgetting to wear the sensors, sensor batteries running out, etc.), 3108 out of 7080 data points (approximately 44%) have the value 0. We represent steps time series data in a matrix where each row represents a person. There are a total of 236 time series objects, each of which has 30 data points, one for each day. Here, we can analyse the long-term usage of pedometers, and the patterns that differentiate the long-term physical performances.

HealthyTogether Dataset. Previously collected for another study (Chen and Pu 2014), this dataset contains the calorie expenditure data of 48 users wearing Fitbit (a wearable accelerometer) for ten days in the period between April 2013 and June 2013. In its raw form, each data point represents activity during a single minute. This dataset do not have any missing values. We process the data in a matrix where each row represents a day. There are a total of 480 time series objects, each of which has 1440 data points. With this dataset, we can analyse the effects of daily routines on the daily physical performance.

Evaluation

Overall Comparison

We compare our method with some well-known baseline algorithms (namely, K-means, 1-nearest neighbor, and agglomerative hierarchical clustering). We employed Euclidean distance for K-means and DTW distance in 1-nearest neighbor and agglomerative hierarchical clustering.

Since the E-Walk and HealthyTogether datasets do not have labels, we evaluate our method via internal cluster evaluation. We specifically employ overall average silhouette width (Kaufman and Rousseeuw 2009). This value indicates the quality of the underlying structure of the clusters: values below 0.25 indicate no structure, values between 0.25 and 0.5 indicate a possibly strong structure, and values above 0.5 indicate a very strong structure (Kaufman and Rousseeuw 2009).

Experiment	Accuracy	F-1	NMI	Jaccard Index
K-means	0.75	0.62	0.51	0.46
Hierarchical	0.81	0.72	0.63	0.58
1-NN	0.93	0.87	0.78	0.77
Our method	0.95	0.92	0.85	0.86

Table 1: The external index scores for the CBF dataset.

Figure 3 conveys the average silhouette width scores for the three datasets. On average, our method outperforms baseline methods in ASW by 0.455, and it is able to capture clusters with high quality. We have applied Wilcoxon signed rank test with $p<0.05$ to compare our method's and baseline methods' ASW scores in each dataset, and validated the significance of these improvements.

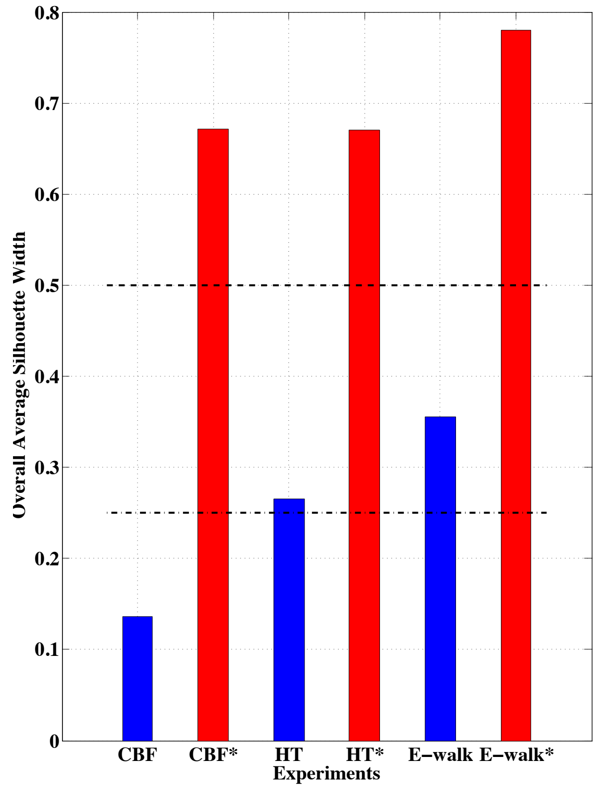

Figure 3: The average silhouette width scores for clustering with (denoted by *) and without our method. "HealthyTogether" is abbreviated as "HT". The lines drawn on 0.25 and 0.5 denote boundary for acceptable and good values of ASW, respectively. We report the highest average score achieved with baseline methods.

Cluster Id	Median of Daily Steps
E1	1842
E2	4194
E3	6461
E4	10357
E5	10646
E6	13782

Table 2: The median of daily step counts for each cluster in E-Walk dataset, with ids matching with those in Figure 4.

CBF Results

Since CBF dataset contains labels, we also evaluated CBF dataset's output clusters with external evaluation indices (accuracy, F-1 score, normalized mutual information and Jaccard index) with 10-fold cross validation. Table 1 summarizes these scores in the CBF dataset. Our approach outperforms baseline methods in terms of accuracy (by 12%), F-1 score (by 0.18), normalized mutual information (by 0.21), and cluster purity (by 0.25). For each of these indices, we

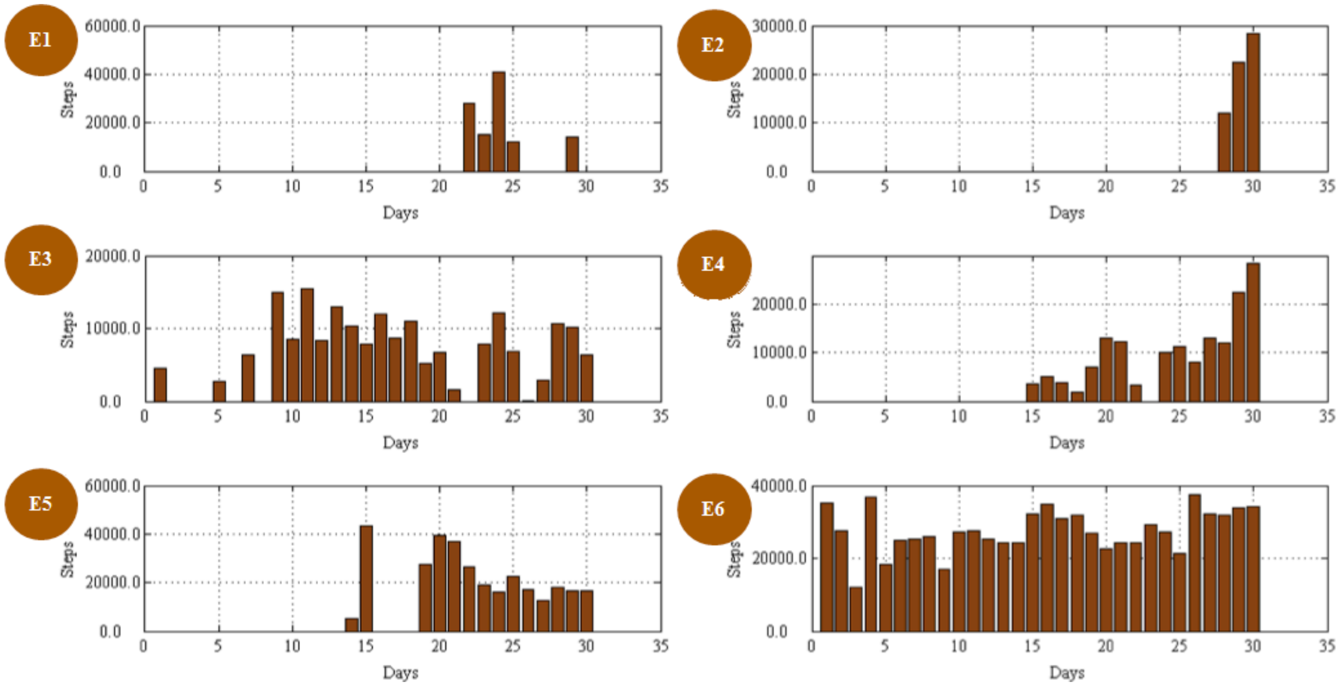

Figure 4: The clusters for the E-Walk dataset ($\lambda = 100$ and $\gamma = 0.065$). Y axis represents the steps taken and X axis represents the days.

compared our method against each of the baseline methods with Wilcoxon signed rank test with $p<0.05$, and validated that these improvements are significant.

E-Walk Results

The representatives for each cluster (member with median number of average steps), and the selected values for the parameters λ and γ are shown in Figure 4. The median calorie expenditures for all clusters are shown in Table 2. Through the 6 clusters that we obtain from this dataset, we can observe the long-term usage patterns of pedometers. For instance, some people convey a novelty effect, i.e., they performed well in the early days of their pedometer usage, but then lost their engagement. Such people are generally grouped in the clusters with lowest average number of steps. We also observe that regularity of activeness has positive contribution towards higher average numbers of steps.

HealthyTogether Results

The representatives for each cluster (member with median calorie expenditure), and the selected values for the parameters λ and γ are shown in Figure 5. The median calorie expenditures for all clusters are shown in Table 3.

The results show that we can characterize 7 types (clusters) of daily activity routines. These routines can be associated to some persona, such as “Commuter” (H1), who has two main peaks in the morning and afternoon; “Afternoon Break-taker” (H2), who is more active in the afternoon with frequent “breaks”; “Early morning person” (H3), who is more active in the early times of the day; “The Frequent breaker” (H4), who takes frequent breaks through the day; “Night Person” (H5), whose is more active late at night; “Hyperactive” (H6), who has moderate, and continuous activeness through the day; and “Traveler” (H7), who has high and continuous activeness through the day.

Through these 7 clusters, we can observe how the intra-day patterns can contribute to the average daily activeness. The average step count increases from the “Commuter” type of daily routine to “Traveler” type of daily routine. Similar to the clustering results in the E-Walk dataset, we see that regular distribution of activeness contributes most to the level of activeness.

Cluster Id	Median of Daily Calories
H1	1412
H2	1519
H3	1587
H4	1640
H5	1660
H6	1862
H7	2353

Table 3: The median of daily step counts for each cluster in HealthyTogether dataset, with ids matching with those in Figure 5.

Conclusion

We proposed a novel approach to perform cluster analysis on ADL data. This approach is different from prior stud-

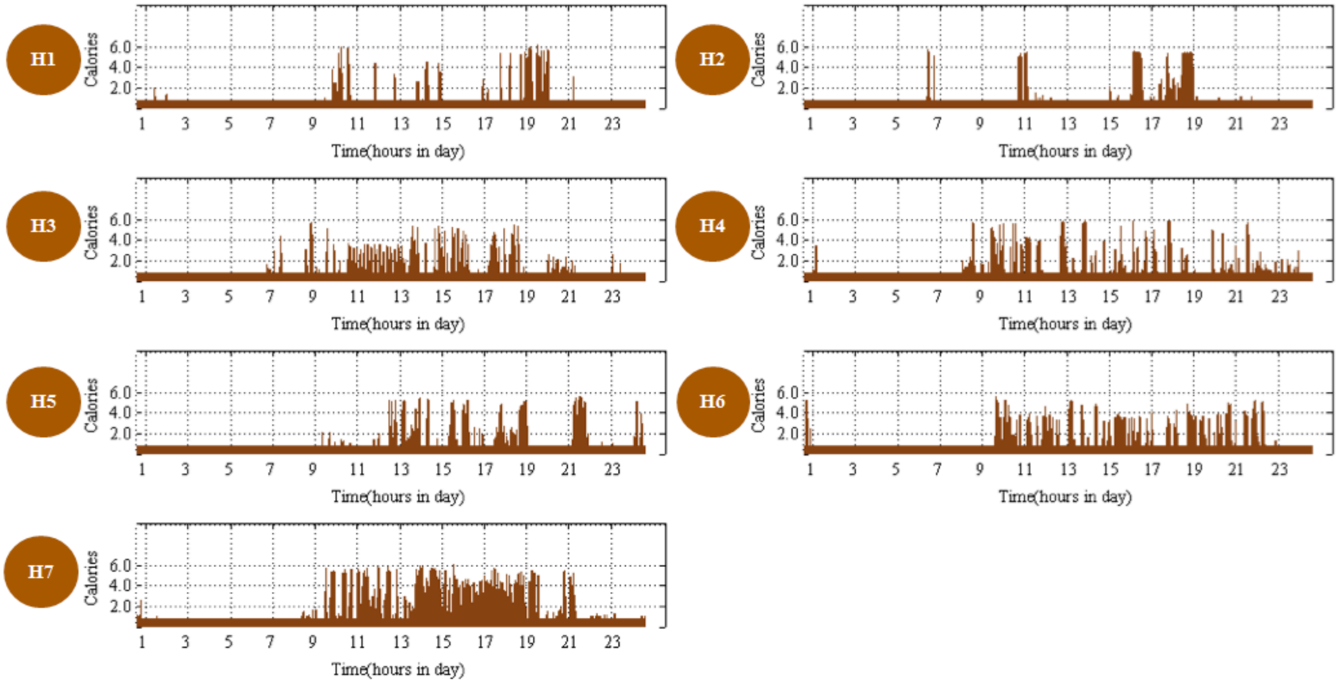

Figure 5: The clusters for the HealthyTogether dataset ($\lambda = 100$ and $\gamma = 0.026$). Y axis represents the calorie expenditure and X-axis represents the hours in the day.

ies as it can process ADL time series without expert knowledge or micro-pattern extraction. Our approach is useful to reveal clusters with high external and internal evaluation scores, and it outperforms baseline algorithms (for instance, by 12% of accuracy and 0.455 points of average silhouette width) with statistical significance. The employed matrix decomposition technique makes our method suitable for high-dimensional data, paving the way for further possible applications such as analysing between-subject variabilities and multi-sensor data.

Our next step is to employ our understandings we obtained from this study to identify and elaborate on predictors or crucial behavior patterns that lend to activeness in daily physical activity routines. Such an analysis of clusters was shown to be useful in predicting illnesses based on behaviour patterns (Madan et al. 2010).

Acknowledgments

We thank the Yiqizou Company (http://yiqizou.com) for providing us with the E-Walk dataset. We thank the Fitbit company (http://www.fitbit.com) for providing us with the access for the time series data of the users of HealthyTogether.

References

Ali, R.; ElHelw, M.; Atallah, L.; Lo, B.; and Yang, G.-Z. 2008. Pattern mining for routine behaviour discovery in pervasive healthcare environments. In *Information Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on*, 241–244. IEEE.

Bao, L., and Intille, S. S. 2004. Activity recognition from user-annotated acceleration data. *Pervasive computing* 1–17.

Berndt, D. J., and Clifford, J. 1994. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, 359–370. Seattle, WA.

Candès, E. J.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM (JACM)* 58(3):11.

Chen, Y., and Pu, P. 2014. Healthytogether: Exploring social incentives for mobile fitness applications. In *The second International Symposium of Chinese CHI 2014, Toronto, Canada, April 26 - 27*. ACM.

Cook, D. J. 2010. Learning setting-generalized activity models for smart spaces. *IEEE intelligent systems* 2010(99):1.

Fu, T.-c. 2011. A review on time series data mining. *Engineering Applications of Artificial Intelligence* 24(1):164–181.

Hodrick, R. J., and Prescott, E. C. 1997. Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking* 1–16.

Kaufman, L., and Rousseeuw, P. J. 2009. *Finding groups in data: an introduction to cluster analysis*, volume 344. Wiley.com.

Keogh, E., and Kasetty, S. 2003. On the need for time series data mining benchmarks: a survey and empirical demonstration. *Data Mining and Knowledge Discovery* 7(4):349–371.

Keogh, E., and Lin, J. 2005. Clustering of time-series subsequences is meaningless: implications for previous and future research. *Knowledge and information systems* 8(2):154–177.

Keogh, E. J., and Pazzani, M. J. 1999. Scaling up dynamic time warping to massive datasets. In *Principles of Data Mining and Knowledge Discovery*. Springer. 1–11.

Kyrillidis, A., and Cevher, V. 2012. Matrix alps: Accelerated low rank and sparse matrix reconstruction. In *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, 185–188. IEEE.

Liang, X.; Ren, X.; Zhang, Z.; and Ma, Y. 2012. Repairing sparse low-rank texture. In *Computer Vision–ECCV 2012*. Springer. 482–495.

Lin, Z.; Chen, M.; and Ma, Y. 2010. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*.

Madan, A.; Cebrian, M.; Lazer, D.; and Pentland, A. 2010. Social sensing for epidemiological behavior change. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, 291–300. ACM.

Min, K.; Zhang, Z.; Wright, J.; and Ma, Y. 2010. Decomposing background topics from keywords by principal component pursuit. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, 269–278. ACM.

Patel, D.; Hsu, W.; and Lee, M. L. 2012. Integrating frequent pattern mining from multiple data domains for classification. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, 1001–1012. IEEE.

Peng, Y.; Ganesh, A.; Wright, J.; Xu, W.; and Ma, Y. 2010. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 763–770. IEEE.

Pham, C.; Plötz, T.; and Olivier, P. 2010. A dynamic time warping approach to real-time activity recognition for food preparation. In *Ambient Intelligence*. Springer. 21–30.

Rashidi, P.; Cook, D. J.; Holder, L. B.; and Schmitter-Edgecombe, M. 2011. Discovering activities to recognize and track in a smart environment. *Knowledge and Data Engineering, IEEE Transactions on* 23(4):527–539.

Rath, T. M., and Manmatha, R. 2003. Word image matching using dynamic time warping. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, II–521. IEEE.

Sakoe, H., and Chiba, S. 1978. Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on* 26(1):43–49.

Sandve, G. K., and Drablos, F. 2006. A survey of motif discovery methods in an integrated framework. *Biol Direct* 1(11).

Smarr, L. 2012. Quantifying your body: A how–to guide from a systems biology perspective. *Biotechnology Journal* 7(8):980–991.

Tollmar, K.; Bentley, F.; and Viedma, C. 2012. Mobile health mashups: Making sense of multiple streams of well-being and contextual data for presentation on a mobile device. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on*, 65–72. IEEE.

Vahdatpour, A.; Amini, N.; and Sarrafzadeh, M. 2009. Toward unsupervised activity discovery using multi-dimensional motif detection in time series. In *IJCAI*, volume 9, 1261–1266.

Wadley, V. G.; Okonkwo, O.; Crowe, M.; and Ross-Meadows, L. A. 2008. Mild cognitive impairment and everyday function: evidence of reduced speed in performing instrumental activities of daily living. *American Journal of Geriatric Psych* 16(5):416–424.

Zheng, J., and Ni, L. M. 2012. An unsupervised framework for sensing individual and cluster behavior patterns from human mobile data. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 153–162. ACM.

Zheng, Y.; Wong, W.-K.; Guan, X.; and Trost, S. 2013. Physical activity recognition from accelerometer data using a multi-scale ensemble method. In *Twenty-Fifth IAAI Conference*.

Zhou, T., and Tao, D. 2011. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 33–40.