A hybrid model for predicting human physical activity status from lifelogging data

Ji Ni [a,c,1], Bowei Chen [b], Nigel M. Allinson [a], Xujiong Ye [a,*]

[a] *School of Computer Science, University of Lincoln, United Kingdom*
[b] *Adam Smith Business School, University of Glasgow, United Kingdom*
[c] *Inception Institute of Artificial Intelligence, United Arab Emirates*

ARTICLE INFO

ABSTRACT

One trend in the recent healthcare transformations is people are encouraged to monitor and manage their health based on their daily diets and physical activity habits. However, much attention of the use of operational research and analytical models in healthcare has been paid to the systematic level such as country or regional policy making or organisational issues. This paper proposes a model concerned with healthcare analytics at the individual level, which can predict human physical activity status from sequential lifelogging data collected from wearable sensors. The model has a two-stage hybrid structure (in short, MOGP-HMM) – a multi-objective genetic programming (MOGP) algorithm in the first stage to reduce the dimensions of lifelogging data and a hidden Markov model (HMM) in the second stage for activity status prediction over time. It can be used as a decision support tool to provide real-time monitoring, statistical analysis and personalized advice to individuals, encouraging positive attitudes towards healthy lifestyles. We validate the model with the real data collected from a group of participants in the UK, and compare it with other popular two-stage hybrid models. Our experimental results show that the MOGP-HMM can achieve comparable performance. To the best of our knowledge, this is the very first study that uses the MOGP in the hybrid two-stage structure for individuals' activity status prediction. It fits seamlessly with the current trend in the UK healthcare transformation of patient empowerment as well as contributing to a strategic development for more efficient and cost-effective provision of healthcare.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

With the significant development of technologies and the radical changes of socio-economic environment, the management planning and decision-making faced by businesses have become more and more complex, requiring the use of sophisticated analytical tools. Operational research techniques (e.g., optimisation, forecasting, simulation) together with other quantitative disciplines (e.g., probability theory, statistics, machine learning, data mining) are particularly useful to solve these challenges (Chen, Kim, Oztekin, & Sundaramoorthi, 2018; Grünig & Kühn, 2013; Hindle & Vidgen, 2018). Therefore, even though the contributions of the above techniques and models themselves are well-documented, the term business analytics has been established over the past decade (Doumpos & Zopounidis, 2016). Business analytics, or simply analytics, uses data, information technology, statistical analysis, mathematical models, optimisation techniques and computer-based simulations to gain improved insight about business operations and make better, fact-based decisions (Evans, 2017). In other words, business analytics is a new multidisciplinary subject which combines the fields of operational research, machine learning, data mining, statistics, big data, and so on Mortenson, Doherty, & Robinson, 2015. It highlights the growing need to use of quantitative approaches for management planning and decision making in a broader context encompassing data, processes, and systems through the integration of traditional problem structuring and solving paradigms with data management and reporting tools, in a way that facilitates learning and action planning in an operational framework (Doumpos & Zopounidis, 2016).

Healthcare is one of the world's largest industries, with many people involved either as employees in healthcare systems or as consumers of healthcare services. Four decades ago, scholars started to use operational research techniques to design healthcare systems and to improve healthcare service delivery (Fries, 1976; Krischer, 1980). The European Working Group on Operational

---

* Corresponding author.
    E-mail addresses: ji.ni@inceptioniai.org (J. Ni), bowei.chen@glasgow.ac.uk (B. Chen), nallinson@lincoln.ac.uk (N.M. Allinson), xye@lincoln.ac.uk (X. Ye).
    [1] Ji Ni is a Senior Applied Scientist at the Inception Institute of Artificial Intelligence. He was a Research Fellow of Machine Learning at the University of Lincoln working on the research of this paper.

Research Applied to Health Services (ORAHS) has been organising annual meetings since 1975. Many of the operational research studies in healthcare have been focused on the application of systematic analysis (Brailsford & Vissers, 2011) such as national or regional policy making and organisational issues. Over the years, technology has revolutionised the way we live, learn and work. It has also been one of the forces driving healthcare transformation. One trend is that people are encouraged to monitor and manage their health based on their daily eating and their physical activity habits based on people-centred healthcare and patient empowerment (World Health Organization, 2014b). For example, Rudner, McDougall, Sailam, Smith, and Sacchetti (2016) reported a case in which a doctor suggested that a patient who had a history of seizures should wear a Fitbit.[2] This device is a wearable sensor that can track the patient's pulse rate and record it through a mobile phone application. The doctor then used the lifelogging data collected from the Fitbit to successfully determine an irregular heart beat that coincided with a grand mal seizure that had occurred three hours earlier. This is a successful application of business analytics in healthcare (sometimes called *healthcare analytics*) at the individual level.

In this paper, we propose a new model concerned with individual healthcare analytcs. Our model can predict human physical activity status from sequential lifelogging data collected from portable devices such as mobile phones and wearable sensors. *Physical activity* refers to any bodily movement produced by skeletal muscles that requires energy expenditure, including activities undertaken while working, playing, travelling, carrying out household tasks and engaging in recreational pursuits (World Health Organization, 2017). According to World Health Organization (2014a), "Insufficient physical activity is one of the 10 leading risk factors for global mortality, causing some 3.2 million deaths each year. In 2010, insufficient physical activity caused 69.3 million disability-adjusted life years (DALYs) – 2.8% of the total – globally". As regular physical activity for adults can reduce the risk of cardiovascular disease, diabetes, cancer and all-cause mortality, the World Health Organization has set a global target to reduce by 10% the prevalence of insufficient physical activity by 2025. Reaching this target requires multisectoral collaboration among government departments and organisations. On an individual level, early disease detection and timely treatment are an effective and economic approach. The use of wearable sensors such as mobile phones, smart watches and fitness trackers to recognise and monitor human activities has recently been investigated for individual health self-management, and it has become an emerging topic in healthcare analytics.

Many conventional studies employ descriptive statistics to summarise lifelogging data and to determine certain thresholds as minimum requirements in terms of daily or weekly walking steps or other metrics to estimate human physical activity status (Caspersen, Powell, & Christensen, 1985; Choi, Pak, Choi, & Choi, 2007; Pate et al., 1995). However, there are two major limitations of those studies. First, human physical activity status in many conventional studies is usually classified into two states, active or inactive, which has limited insights and prevents broader applications. Fine-grained classification can be further investigated to measure physical activity status. The second limitation is that many conventional studies only illustrate the static characteristics of data without considering historical information. This limitation is particularly evident in the case of individual health self-management. The pattern of physical activity from one person to the next is different. Therefore, when high dimensional sequential lifelogging data is collected from wearable sensors,

it is worth considering individuals' sequential activities and the effects of previous activities on the current activity status (Gurrin, Smeaton, & Doherty, 2014; Zhou & Gurrin, 2012).

Our proposed model has a two-stage hybrid structure (in short, MOGP-HMM). It contains a multi-objective genetic programming (MOGP) algorithm in the first stage and a hidden Markov model (HMM) in the second stage. The MOGP alleviates the first limitation mentioned above. It is a multi-class classifier that transforms a high-dimensional feature space of the collected lifelogging data into a new discrete class space which represents activity observation. The HMM in the second stage addresses the second limitation. It is a chain-structured Bayesian network which can be used to exploit the sequential patterns from observations. Simply put, an individual's physical activity status at a time is described by a latent variable. Latent variables over time are connected through a Markov process rather than being independent of each other. Since scoring systems have been widely used in assessing quality of life (QoL) such as QoL questionnaire VF-14 (Terwee, Gerding, Dekker, Prummel, & Wiersinga, 1998) and SF-12 (Gandek et al., 1998), observation and physical activity status in our study are both expressed in terms of a measurement score ranging from the inactive state to the highly active state. Given a time series of observations, the HMM can predict an individual's activity status accordingly. We validate the model with the real lifelogging data collected from a group of participants in the UK, and conduct experiments in a supervised learning setting (Bishop, 2007) where the scores (or states) of activity status are labelled based on the UK national health guidelines (UK National Health Service, 2015). We also compare our model with another popular hybrid model SVM-HMM which combines a support vector machine (SVM) with a HMM. Our experimental results show that the MOGP-HMM can achieve comparable performance as the SVM-HMM. However, Unlike SVMs, our MOGP-HMM model is not sensitive to the choice of kernel functions and thus provides more robust and discriminative representations of sparse data.

The research of this paper is multidisciplinary, which contributes to the recent use of operational research, machine learning, data mining, big data and the Internet of things in healthcare analytics. Firstly, this is one of the few studies which discuss the implementation of operational research in healthcare at the individual level (Royston, 1998). In the meantime, lifelogging data is truly a big data problem because it is multidimensional, it contains many different features in terms of different formats, and it can be retrieved continuously from wearable sensors. We develop a two-stage model to reduce the complexity of lifelogging data and then to predict an individual's physical activity status over time. In essence, the proposed model is a personalized data-driven model based on the state-of-the-art machine learning algorithms so it contributes to the applications of machine learning. Further, our model can be deployed on a cloud server and can be used as a decision support tool to provide real-time monitoring, statistical analysis and personalized advice to an individual through portable digital devices. Therefore, it can be a practical application of the Internet of things in healthcare. Within the field of business analytics, our proposed model contains technology, quantitative methods and decision making. As indicated by Mortenson et al. (2015), they are the key elements of business analytics. Similar to the existing studies (Dag, Oztekin, Yucel, Bulur, & Megahed, 2017; Dag, Topuz, Oztekin, Bulur, & Megahed, 2016; Harris, May, & Vargas, 2016; Roumani, Roumani, Nwankpa, & Tanniru, 2018; Topuz, Uner, Oztekin, & Yildirim, 2018), our proposed model deals with predictive analytics. From a high-level perspective in healthcare, this study fits seamlessly with the current trend in the UK healthcare for patient empowerment, and contributes to a strategic development for the provision of more efficient and cost-effective healthcare.

---

[2] https://www.fitbit.com.

Technology wise, using the MOGP also provides methodological contributions in the two-stage hybrid modelling for physical activity prediction. It is a non-parametric optimisation classifier, differing from many genetic algorithms and machine learning models where parameters need to be set or trained in advance. It uses Pareto dominance to optimally select GP tree models considering the trade-off between the model fitness and complexity. Therefore, the MOGP is more efficient and robust. Unlike the SVM, it is not sensitive to the choice of kernel functions and thus provides more robust and discriminative representation of sparse data. As lifelogging data is usually sparse and noisy due to the fact that each individual usually has his or her own activity pattern, the MOGP algorithm seems more suitable than the SVM in activity learning. Although GP algorithms have been used to evolve probabilistic trees that search for the optimal topology in bioinformatics (Won, Hamelryck, Prügel-Bennett, & Krogh, 2007) and stock trading (Chen, Mabu, Shimada, & Hirasawa, 2009; Ghaddar, Sakr, & Asiedu, 2016), to the best of our knowledge, this is the first work that a MOGP algorithm has been used as a multi-class classifier to construct a classification-HMM hybrid model for solving sequential learning problems. Our model can be of interest and easily adapted to other relevant domains in business analytics, such as consumer choice modelling (Blanchet, Gallego, & Goyal, 2016; Sandıkcı, Maillart, Schaefer, Alagoz, & Robert, 2008) and high dimensional business data classification or dimension reduction (Debaere, Coussement, & De Ruyckc, 2018; Ghaddar & Naoum-Sawaya, 2018).

The remainder of the paper is organised as follows. Section 2 reviews the related literature. Section 3 introduces our proposed hybrid model. Section 4 describes our data, presents experimental results and gives an analysis. Section 5 concludes the paper.

2. Related work

Our study touches upon several streams of literature. In the following discussion, we review the related work in both healthcare and hybrid learning machines. For the former, we first discuss the recent studies on the use of operational research in healthcare at the country and organisational levels, and then individual health monitoring, prediction and self-management using wearable sensors. For the latter, we discuss the basic concepts and settings of hybrid learning machines and compare the related two-stage hybrid models.

Operational research has been used and developed for healthcare over the years in the hope of improving the healthcare effectiveness and efficiency as well as controlling or reducing the costs (Brailsford & Vissers, 2011; Fries, 1976; Krischer, 1980). A significant proportion of earlier studies has examined healthcare systems at the country or organisational level (Brailsford & Vissers, 2008; Kunc, Harper, & Katsikopoulos, 2018), such as national healthcare policy making or management, organisational issues and service delivery. For example, at the national level, Hindle, Hindle, and Hindle (2013) proposed a decision support framework based on geographical modelling for the strategic management of radical changes in hospital services in Northern Ireland. Denoyel, Alfandari, and Thiele (2017) designed a structured optimisation model for bill payers combining reference pricing and tiered network for novel healthcare payment policies in the United States. Willis, Cave, and Kunc (2018) proposed a multi-methodology approach for healthcare workforce planning in England. Hejazi, Badri, and Yang (2019) discussed a reliability-based approach to measure healthcare system performance for policy makers. At the organisational level, Tako and Kotiadis (2015) proposed a framework to support facilitated simulation modelling in healthcare. Li, Vo, Randhawa, and Fick (2017) designed utilization-based spatial accessibility decision support systems for patients.

Kunc et al. (2018) further investigated the importance of human behaviour aspects in the application of operational research in healthcare at an organisational level by reviewing 130 related papers. Rouyendegh, Oztekin, Ekong, and Dag (2018) proposed a data envelopment analysis based fuzzy multi-criteria decision making model to enhance the business performance of companies in the healthcare industry.

Royston (1998) pointed out that prevention and treatment based on each patient's knowledge and habit are part of the key shift patterns of using the operational research in healthcare for the 21st century. Our study in this paper is concerned with healthcare analytics at the individual level. Specifically, we are focused on individual health monitoring, prediction and self-management using wearable sensors. It should be noted that wearable sensors here refer to mobile phones, smart watches, fitness trackers, and ad-hoc wearable devices like Shimmer.[3] There are two groups of related literature. The first group analyses the vital signs provided by wearable sensors (Banaee, Ahmed, & Loutfi, 2013) such as electrocardiogram, oxygen saturation, heart rate, photoplethysmography, blood glucose, blood pressure and respiratory rate. The second group is focused on recognising and monitoring individual human activities (Kulev, Pu, & Faltings, 2016; Liao, Fox, & Kautz, 2005; Luque, Casilari, Morón, & Redondo, 2014; Micucci, Mobilio, & Napoletano, 2017; Vilarinho et al., 2015), which also overlaps with the fields of computer vision, machine learning and data mining. Our study in this paper is closer to the second group. We use wearable sensors to collect lifelogging data from a group of participants in the UK. Further details about our data are discussed in Section 4. It is worth mentioning the following two studies in the second group. Liao et al. (2005) discussed a general framework for activity recognition by building upon and extending relational Markov networks. The model includes a variety of features including temporal information, spatial information and global constraints, so human activity locations (e.g., home, work, shop, dinning, etc.) can then be predicted. Kulev et al. (2016) proposed a mixture model to understand how the intervention affects daily human activities, whether they increase or decrease the amount of physical activities at each moment during the day. Two types of information are relevant: the person's daily activity pattern before the intervention and their activity change pattern after the intervention. The model is used to find the latent structure in a heterogeneous population.

Hybrid models have been widely used in machine learning to solve different real world problems. In some reference, they are called *hybrid learning machines* (Abraham, Corchado, & Corchado, 2009) or *intelligent hybrid systems* (Goonatilake & Khebbal, 1995). As hybrid learning machines use different types of models, here we explain some important concepts and theories. According to Domingos (2015), there are five major tribes in machine learning or artificial intelligence in general: symbolists, connectiontists, evolutionaires, Bayesians and analogizers. Symbolists believe all intelligence can be reduced to manipulating symbols and they solve problems using pre-existing knowledge. Many expert systems use the symbolists' approaches to solve problems with a set of rules (Zhang & Zhang, 2014) and fuzzy logic is the attempt of symbolists at tackling uncertainties (Zadeh, 1965). Connectionists hope to use artificial neural networks to represent manmmalian neural systems such as deep neural networks (Goodfellow, Bengio, & Courville, 2016). Evolutionaires are influenced by Darwin's theory on evolution and believe that all learning arises from natural selection such as genetic programming (Koza, 1992). Bayesians are concerned above all with uncertainty and their theories are heavily based on probabilistic inference and Bayes' theorem such as the HMM (Bishop, 2007). Analogizers are the least cohesive of

---

the five tribes (Domingos, 2015), recognising similarities between situations and thereby inferring other similarities such as the SVM (Vapnik, 2000). A hybrid learning machine can contain at least two machine learning models from one tribe or different tribes. It could be called the *hybrid neural system* (Wermter & Sun, 2000) if all models come from the connectionists' tribe such as the work of Borrajo, Baruque, Ccorchado, Bajo, and Corchado (2011). The models in a hybrid learning machine can be used in parallel or by sequence. For example, Peddabachigari, Abraham, Grosan, and Thomas (2007) discussed a hybrid learning machine combining a decision tree and an SVM for intrusion detection. The system takes prediction outputs from two models (for example, votes) and then combines them into the final output. This is also called *ensemble learning* (Zhou, 2012), in which multiple models (called *base learners*) are strategically combined to create a stronger model to solve a particular problem. De Caigny, Coussement, and De Bock (2018) designed a hybrid model based on a decision tree in the first stage and a logistic regression in the second stage. The output of the decision tree is the input of the logistic regression, the output of which is the system's final output. In this paper, our proposed hybrid model MOGP-HMM contains two models used by sequence from two tribes (i.e., the MOGP is from evolutionaires and the HMM is from Bayesians). We also compare it with the benchmarked SVM-HMM.

From the functional perspective, our proposed hybrid model MOGP-HMM and the benchmark SVM-HMM can be expressed as the classification-HMM. The MOGP or SVM is used for classification in the first stage while the second stage HMM is used in finding and modelling patterns in sequential data, satisfying Markovian property. In the previous studies, SVMs with different kernels and the Gaussian mixture model (GMM) have been used in the classification-HMM structure. For example, the SVM-HMM has been successfully applied in speech recognition (Mohameda & Nair, 2012; Stadermann & Rigoll, 2004), metadata extraction (Zhang, Yin, Deng, & Yang, 2008), and vision based human behaviour recognition (Han, Zhang, Li, & Yao, 2014). The GMM-HMM has been used in vision based human motion detection (Concha, Yi, Xu, Moghaddam, & Piccardi, 2011; Han et al., 2014). According to Han et al. (2014), the SVM-HMM achieves a better recognition performance than the GMM-HMM in short video sequences because the SVM can clearly distinguish the differences between categories in consecutive frames. Although the SVM has shown great success in the previous studies, it has several limitations. First, choosing an appropriate kernel function is always a challenging task as it requires cross validation and it is data and task dependent (Auria & Moro, 2008). Second, the SVM usually needs a long training time for large datasets. In this study we aim to find an alternative classifier which is efficient as well as robust. The MOGP has achieved a wide range of success inclusive of applications to classification problems (Ni & Rockett, 2014; Shao, Liu, & Li, 2014; Zhang & Rockett, 2009) but it has not been used in the classification-HMM hybrid structure for healthcare applications. Similar to the SVM, the MOGP is a non-parametric model which requires fewer assumptions about the data, and consequently performs better in situations where the true distribution is unknown. However, the modelling process of the MOGP is totally different to the SVM because it is from the evolutionaires' tribe. In essence, the MOGP is a tree-based algorithm, which can provide a better visualisation graph on the solution. Also, unlike SVMs, it is not sensitive to the choice of kernel functions and thus provides more robust and discriminative representations of sparse data. The evolutionary process searches for a richer model space to minimise both 0/1 loss and the size of decision trees using Pareto dominance (Poli, Langdon, & McPhee, 2008). Therefore, in this paper we use the MOGP. Apart from the theoretical comparison between the SVM and the MOGP here, we also empirically compare the MOGP-HMM
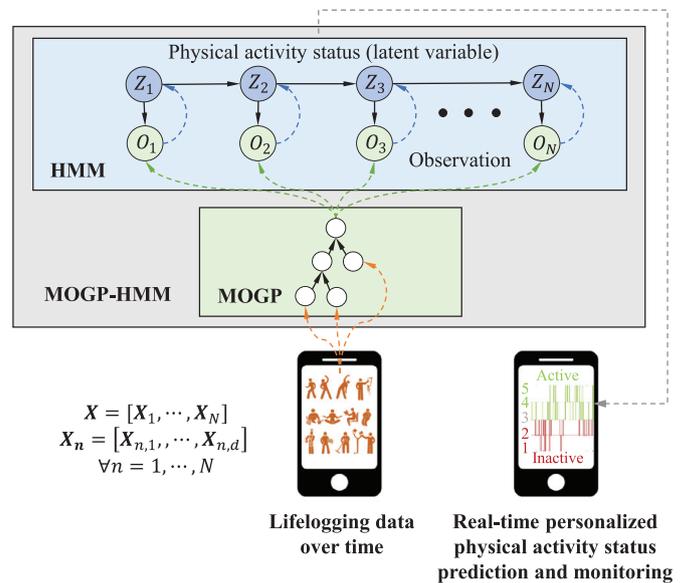
Fig. 1. Schematic view of the MOGP-HMM. The dotted lines show how the collected sequential lifelogging data is processed to predict fine-grained human physical activity status. The model can be deployed on a cloud server and provides real-time monitoring and personalized health advice to an individual through portable digital devices.

and the SVM-HMM (with different kernels) based on our data in Section 4.

3. The MOGP-HMM

The proposed MOGP-HMM contains two stages: (i) a MOGP algorithm in the first stage; and (ii) a first-order HMM in the second stage. Fig. 1 presents a schematic view of the MOGP-HMM. The first-order HMM is represented as a chain-structured Bayesian network where $Z_1, \ldots, Z_N$ are the latent variables representing the human physical activity status over a finite time horizon $t_1, \ldots, t_N$. and $O_1, \ldots, O_N$ are the observations obtained by the MOGP algorithm based on the collected lifelogging data $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_N]$ where $\mathbf{X}_n = [\mathbf{X}_{n,1}, \ldots, \mathbf{X}_{n,d}]$ for $n = 1, \ldots, N$ and d is the dimension of the feature space.

3.1. Classifying lifelogging data using the MOGP algorithm

In the first stage, the MOGP takes an input vector $\mathbf{X}_n$ of an individual's lifelogging data at time t_n and assigns it to one of M discrete classes representing observation states. To simplify the notation, the observation O_n takes a value from a set of integers $S_O = \{1, \ldots, M\}$. In the following discussion, we explain what a GP tree is and we show how it works as a binary classifier. We then introduce how GP trees are built and how the optimal tree models are determined under multiple objectives. Finally, we discuss the ensemble method used to create a multi-class classifier.

GP algorithms use tree-based syntax to present a function $f(\cdot)$ which can transform an input vector $\mathbf{X}_n = [\mathbf{X}_{n,1}, \ldots, \mathbf{X}_{n,d}] \in \mathbb{R}^d$ from a d-dimensional feature space into a 1-dimensional decision space $Y \in \mathbb{R}$, where the *leaf nodes* take the input vector, the *internal nodes* specify the arithmetic operations and the *root node* gives the response in the decision space. Therefore, GP trees can be used to solve binary classification problems. Fig. 2 presents a toy example of a GP tree, in which each input vector has three features (i.e., $\mathbf{X}_n = [\mathbf{X}_{n,1}, \mathbf{X}_{n,2}, \mathbf{X}_{n,3}]$) and the GP tree function then gives a response $y_n = f(\mathbf{X}_n) = (\mathbf{X}_{n,1} + \mathbf{X}_{n,3}) \times \mathbf{X}_{n,2}$. If the training lifelogging data has $\tilde{N}$ input vectors, then $\tilde{N}$ responses can be obtained in the decision space. Therefore, an optimal response can be found from
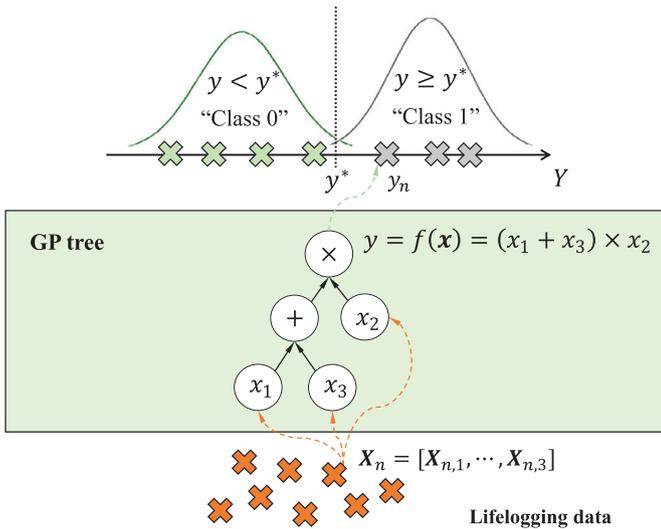
Fig. 2. Illustration of using a GP tree as a binary classifier. The collected lifelogging data is highlighted in orange. The GP tree projects the multidimensional lifelogging data into one dimensional space Y where the classification is based on the threshold value y^* and the classified data is highlighted by green and blue colour, respectively.

Table 1
Experimental settings for training the MOGP algorithm.

Description	Setting
Population size	100
Initialization	Ramped half-and-half method (Koza, 1992)
Termination criterion	0/1 loss = 0 or 80,000 evaluations
Crossover and mutation	Point crossover (Koza, 1992)
	Point mutation (Koza, 1992)
	Tree depth = 4
Node type	Unary minus
	Addition
	Subtraction
	Multiplication
	Analytic quotient (Ni et al., 2013)

the set of obtained $\widetilde{N}$ responses and be used as the threshold y^* to classify inputs so that the misclassification error e^* is minimised, as illustrated in Algorithm 1.

Algorithm 1 Searching for the threshold y^* in a GP tree.

1: **Input: X, L,** $f(\cdot)$ ▷ Lifelogging data, label, GP function
2: **for** $n = 1, \ldots, \widetilde{N}$ **do**   ▷ $\widetilde{N}$ instances
3:     $y_n \leftarrow f(\mathbf{X}_n)$ ▷ $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_{\widetilde{N}}]$ is a $\widetilde{N} \times d$ matrix
4: **end for**
5: **for** $n = 1, \ldots, \widetilde{N}$ **do**
6:     $y_n^* \leftarrow y_n$
7:     **for** $\tilde{n} = 1, \ldots, \widetilde{N}$ **do**
8:        $\mathbf{D}_{n,\tilde{n}} \leftarrow \mathbb{I}_{\{y_{\tilde{n}} > y_n^*\}}$
9:     **end for**
10:    $e_n \leftarrow \frac{1}{\widetilde{N}} \sum_{\tilde{n}=1}^{\widetilde{N}} \mathbb{I}_{\{\mathbf{D}_{n,\tilde{n}} \neq l_{\tilde{n}}\}}$ ▷ $\mathbf{L} = [l_1, \ldots, l_{\widetilde{N}}]$ is a $\widetilde{N} \times 1$ vector
11: **end for**
12: $e^* \leftarrow \min\{e_1, \ldots, e_{\tilde{N}}\}$; $n^* \leftarrow \operatorname{argmin}_{\{1,\ldots,\tilde{N}\}}\{e_1, \ldots, e_{\tilde{N}}\}$; $y^* \leftarrow y_{n^*}^*$
13: **Output:** (y^*, e^*)

Similar to other evolutionary algorithms, the individuals in the initial population are randomly generated in GP algorithms. Here we adopt the widely used Ramped half-and-half method (Koza, 1992), which generates a full sub-tree on one half of the root and a random tree with various size and shapes on the other. The example tree shown in Fig. 1 is the case where the left half is a full tree while the right half is not. We also use point crossover and mutation, as illustrated in Fig. 3. Given two parents, point crossover randomly selects a crossover point in each parent tree. It then creates the offspring by replacing the sub-tree rooted at the crossover point in a copy of the first parent with a copy of the sub-tree rooted at the crossover point in the second parent. Point mutation randomly selects a mutation point in a tree and substitutes the sub-tree rooted there with a randomly generated sub-tree. More details about our experimental settings of GP trees are summarised in Table 1 in Section 4.

In the evolutionary process, a GP algorithm searches for the *global optima* of the specified objective function. If the misclassification error is set as the only objective, the finally selected tree model may fit the training data excessively and end up overfitting. In many machine learning and data mining techniques,

regularization is added to avoid overfitting (Bishop, 2007). However, this will increase the training efforts. In the paper, we use an alternative method to reduce overfitting. The tree size (i.e., the number of tree nodes) is set as the second objective in the optimisation. This can preserve simpler models and improve model generalisation. Pareto dominance is employed to compare and rank vectors of multiple objectives. Let $\mathbf{P} = [p_1, \ldots, p_W]$ and $\mathbf{Q} = [q_1, \ldots, q_W]$ be two W-dimensional vectors. Mathematically, $\mathbf{P}$ is said to (Pareto) *dominate* $\mathbf{Q}$, denoted by $\mathbf{P} \prec \mathbf{Q}$, if the following two conditions are satisfied:

$$p_w \leq q_w, \quad \forall w \in \{1, \ldots, W\}, \tag{1}$$

$$p_w < q_w, \quad \exists w \in \{1, \ldots, W\}. \tag{2}$$

In our optimal selection, the highest rank 1 is assigned to a tree if there are no other trees that dominate it. Trees which are not dominated by the rank 1 tree are then assigned to rank 1. We exclude all rank 1 trees and repeat the procedure to assign rank 2 to trees which are dominated by each other. Then rank 2 trees are excluded and the procedure is repeated until all tree models are assigned a rank. For example, we have five 2-dimensional objective vectors presenting the misclassification error and node count of five different trees: (0.213, 28), (0.213, 67), (0.197, 85), (0.322, 15), (0.225, 50). In the first round, (0.213, 28) dominates (0.213, 67) and (0.225, 50). However, (0.213, 28), (0.197, 85), (0.322, 15) do not dominate each other, so they are assigned to rank 1 and they form the *Pareto frontier*. In the second round, (0.213, 67) and (0.225, 50) do not dominate each other so both are assigned to rank 2. The advantage of using the multi-objective optimisation imposes the simplicity of the models as a form of regularisation in the optimisation procedure and improves the model generalisation.

The MOGP algorithm discussed so far is a binary classifier. To solve multi-class classification problems, an ensemble method is used to merge a number of binary classifiers. Specifically, as illustrated in Fig. 4, if there are K classes (or states) labelled in the lifelogging data, for a class $k = 1, \ldots, K$, a MOGP tree can be obtained from the Pareto frontier with respect to the binary classification problem of 'Class k' or 'Non-class k'. Therefore, K MOGP trees can be obtained, and they can be sorted in ascending order based on their misclassification errors, denoted by $f_{(1)}, \ldots, f_{(K)}$. It should be noted that the notation (k), $k = 1, \ldots, K$, represents the index of the sorted tree model but not the class that the tree solves. For example, $f_{(1)}$ can be the tree model that classifies data into 'Class 2' or 'Non-class 2'. We start with $f_{(1)}$ and classify the training data into either 'Class (1)' or 'Non-class (1)'. The training data of the former is excluded and the rest of data is then classified by $f_{(2)}$. This step is repeated until $f_{(K)}$ and the data of 'Non-class (K)' is assigned to 'Class 0'. Therefore, lifelogging data can be classified into $M = K + 1$ classes (or states) of observations. An advantage of our method is that an additional class is created in the HMM observations. This gives a finer classification of lifelogging data as well
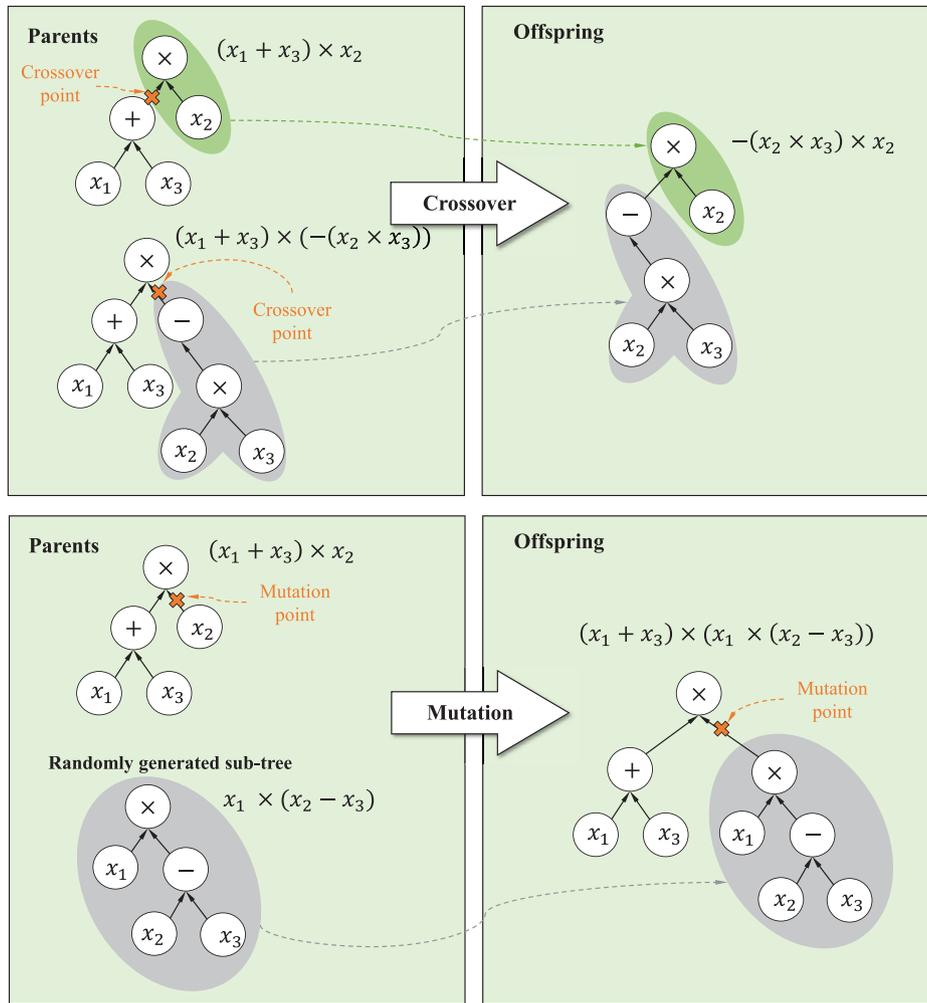
Fig. 3. Illustration of point crossover and mutation.
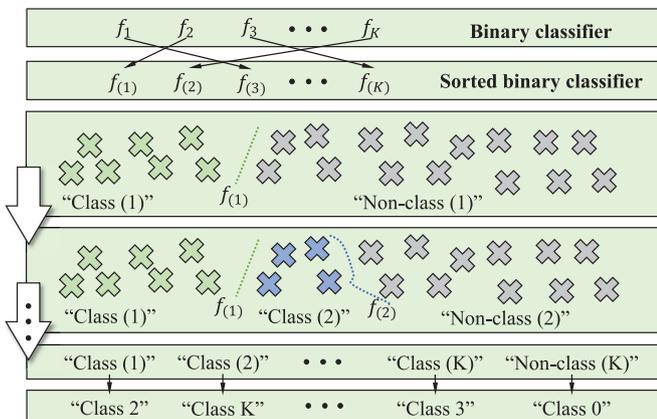
Fig. 4. Illustration of creating a multi-class MOGP classifier. $f_1, \ldots, f_K$ are binary MOGP tree models based on the labelled classes (or states); $f_{(1)}, \ldots, f_{(K)}$ are the sorted binary MOGP tree models in ascending order based on their misclassification errors.

as avoids the case that the one-to-one mapping of the classes of observations and latent variables in the HMM in the second stage.

3.2. Predicting physical activity status using the HMM

In the second stage, we use a first-order HMM (Bishop, 2007; Ghahramani, 2001) to predict an individual's physical activity

status when a sequence of observations is given. As illustrated in Fig. 1, an individual's physical activity status at time t_n is described by the latent variable Z_n, which takes the value from a set of integers $\mathcal{S}_Z = \{1, \ldots, K\}$. Latent variables are connected through a first-order Markov chain in which the distribution $\mathbb{P}(Z_n \mid Z_{n-1})$ of Z_n is conditioned on the value of the previous value Z_{n-1}. Since there are K states, this conditional distribution corresponds to a $K \times K$ matrix that we denote by $\mathbf{A}$, the elements of which are known as *transition probabilities*, i.e, $\mathbf{A}_{i,j} = \mathbb{P}(Z_n = j \mid Z_{n-1} = i)$ where $i, j \in \mathcal{S}_Z$. Latent variables are not observed directly. However, each latent variable Z_n determines an observation O_n through the conditional distribution $\mathbb{P}(O_n \mid Z_n)$. As there are M classes of observations, this conditional distribution corresponds to a $K \times M$ matrix $\mathbf{B}$ whose elements are called *emission probabilities*, i.e., $\mathbf{B}_{i,j} = \mathbb{P}(O_n = j \mid Z_n = i)$ where $i \in \mathcal{S}_Z, j \in \mathcal{S}_O$. Therefore, the following joint distribution can express the relationship among a sequence of observations:

$$\mathbb{P}(Z_{1:N}, O_{1:N}) = \mathbb{P}(Z_1) \left(\prod_{n=2}^{N} \mathbb{P}(Z_n \mid Z_{n-1}) \right) \left(\prod_{n=1}^{N} \mathbb{P}(O_n \mid Z_n) \right), \quad (3)$$

where $Z_{1:N}$ represents $Z_1, \ldots, Z_N$, and $\mathbb{P}(Z_1)$ is the initial latent state probability. As there are K states of the latent variable, the initial latent state probability can be denoted by a $K \times 1$ vector $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_K]$.

The model parameters $\{\boldsymbol{\pi}, \mathbf{A}, \mathbf{B}\}$ can be estimated using the Baum-Welch algorithm (Bishop, 2007). It is essentially an expectation-maximization (EM) algorithm that estimates the values of parameters to maximize $\mathbb{P}(O_{1:N}; \boldsymbol{\pi}, \mathbf{A}, \mathbf{B})$. However, the
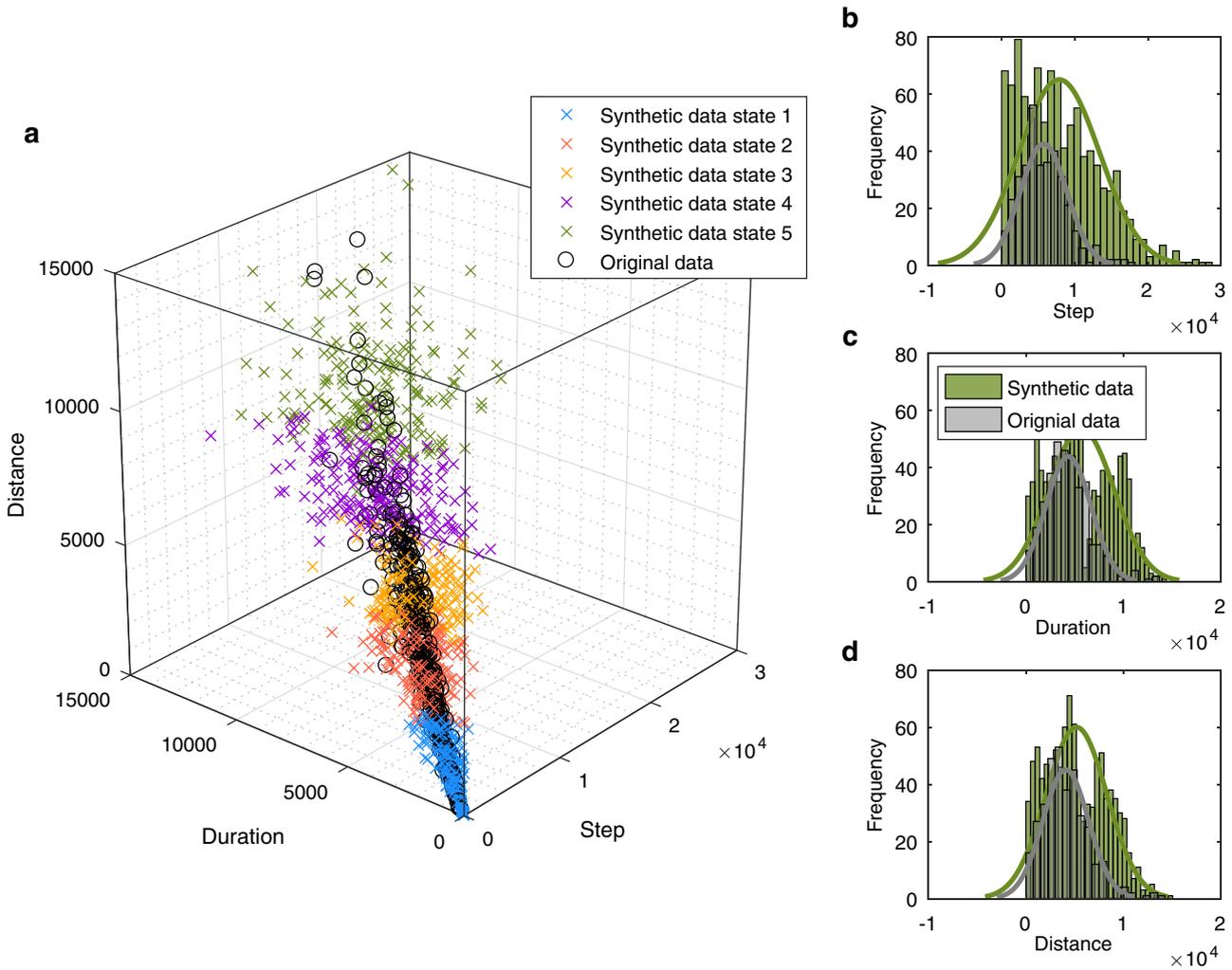
Fig. 5. Example of the original real data and the synthetic data for a participant: (a) the combination of both data and the tagged classes/states in the synthetic data are highlighted by different colours; (b) the histogram and the fitted Gaussian density of both original and synthetic data for the variable step; (c) the histogram and the fitted Gaussian density of both original and synthetic data for the variable duration; (d) the histogram and the fitted Gaussian density of both original and synthetic data for the variable distance.

accuracy of the estimate varies. As the observations are obtained from the training data in the first stage and the training data has been labelled, the model parameters can be estimated based on the ground truth as follows:

$$\pi_i = \frac{\#(Z_n = i)}{\widetilde{N}}, \quad i \in \mathcal{S}_Z, \tag{4}$$

$$\mathbf{A}_{i,j} = \frac{\#(Z_{n-1} = i, Z_n = j)}{\#(Z_{n-1} = i)}, \quad i, j \in \mathcal{S}_Z, \tag{5}$$

$$\mathbf{B}_{i,j} = \frac{\#(Z_n = i, O_n = j)}{\#(Z_n = i)}, \quad i \in \mathcal{S}_Z, j \in \mathcal{S}_O, \tag{6}$$

where the notation # counts the occurrence number and $\widetilde{N}$ is the size of the training data, e.g., the initial latent state probability π_i is equal to the number of occurrences of state i divided by the size of the training data.

Given observations $O_{1,N}$ and the model $\{\pi, \mathbf{A}, \mathbf{B}\}$, how do we find the latent variable sequence $Z_{1:N}$ that best represents the observations? This corresponds to finding the most probable sequence of latent variable states, and this can be solved efficiently using the Viterbi algorithm (Bishop, 2007). Simply put, the most probable latent variable state at time t_N can be obtained

by $Z_N^* = \operatorname{argmax}_{i \in \mathcal{S}_Z} \delta_N(i)$, where $\delta_n(i) \triangleq \max_{Z_{1:(n-1)}} \mathbb{P}(Z_{1:(n-1)}, Z_n = i \mid O_{1:n}), \forall n = 1, \ldots, N$, and the most probable sequence can be computed using traceback.

4. Experiments

In this section, we introduce the collected lifelogging data, describe our experimental settings, and give an analysis of the experimental results.

4.1. Data

Our lifelogging data was collected through the Moves mobile application,[4] which uses accelerometer and GPS sensors in a mobile phone to automatically record any walking, cycling, and running activities of its user. It contains the activities recorded from 10 different participants in the UK, ranging from 118 to 401 days. For each activity, the variables (or features) steps, distance, and duration are collected. Based on the UK national health guidance (UK National Health Service, 2015), the physical activity status is explicitly labelled as a measurement score ranging from 1 (inactive

---

[4] http://www.moves-app.com.
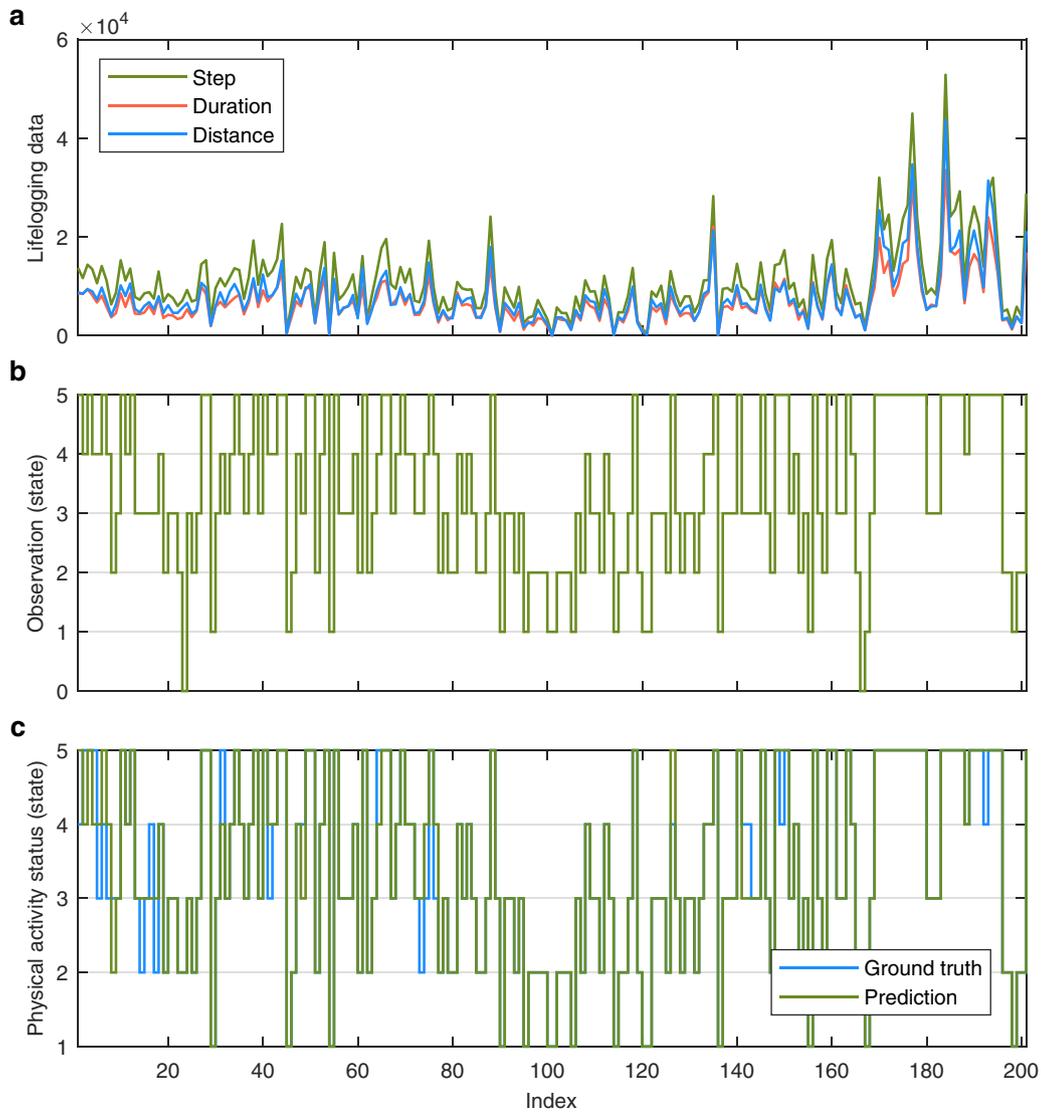
Fig. 6. Example of using the MOGP-HMM for a participant: (a) the time series plot of real lifelogging data; (b) the stairstep plot of observations obtained by the MOGP algorithm; (c) the stairstep plot of the predicted and labelled physical activity statuses, respectively.

state) to 5 (active state). However, the behavioural characteristics vary from user to user – some people live an inactive life in which a highly active pattern is rarely observed while some people moves a lot every day. To overcome the problem that the states of an individual's physical activity status are imbalanced, we generate synthetic data and use them together with the original real data in the experiments. Specifically, for each participant, the synthetic data is only used for training the MOGP algorithm in the first stage, and the original real data is used in the second stage for estimating the HMM and predicting the user's physical activity status over time.

The following strategy is used to create the synthetic data. Two intermediate variables are defined $H = Distance/Duration$ and $R = Steps/Duration$. Their sample mean and standard deviations can be obtained from the raw data. A new duration value can be sampled from the raw data, which can be multiplied by $\mathcal{N}(\mu_H, \sigma_H^2)$ and $\mathcal{N}(\mu_R, \sigma_R^2)$ to create the values of corresponding distance and step, where $\mathcal{N}$ represents Gaussian distribution. It should be noted that the generate values are truncated to be non-negative numbers. Fig. 5 presents an example of the synthetic data and the original raw data for a participant. The classes of physical activity status

can be clearly identified and each class has a certain amount of data. The histogram and the fitted Gaussian density plots of both original and synthetic data for the input variables exhibit similar and consistent distributions. It should be noted that our study is limited to the observations in the real data. Gaussian or Gaussian-like distribution is simple and can specify both central tendency and dispersion of data with parameters mean and standard deviation. The used left-side truncated Gaussian distribution (Burkardt, 2014) is a popular parametric method used to generate synthetic data when there is a lack of real data for training models.

In order to evaluate the robustness of the proposed hybrid model, white noise is generated at different levels ranging from 0 to 0.2 and is incorporated into the labelling process of lifelogging data. Specifically, we follow the UK medical guidance (UK National Health Service, 2015) to label data and add a noise term into the variables step, distance, and duration based on their standard deviations, respectively. Therefore, slightly different labels, i.e., the states of human physical activity status, are obtained under different noise settings. This takes into consideration that doctors may have slightly different ratings for a participant's physical activity status.

Table 2

Test errors (i.e., the numbers in brackets) and relative rankings based on test errors of four hybrid models under different noise settings for one participant.

Noise	MOGP-HMM	SVM(R)-HMM	SVM(P)-HMM	SVM(S)-HMM
0	2 (0.0119)	1 (0.0085)	3 (0.0261)	4 (0.0283)
0.01	1 (0.0145)	2 (0.0167)	3 (0.0255)	4 (0.0308)
0.02	1 (0.0205)	2 (0.0227)	3 (0.0308)	4 (0.0368)
0.03	2 (0.0419)	3 (0.0425)	1 (0.0324)	4 (0.0453)
0.04	3 (0.0573)	2 (0.0567)	1 (0.0463)	4 (0.0595)
0.05	2 (0.0607)	3 (0.0652)	1 (0.0548)	4 (0.0680)
0.06	2 (0.0727)	4 (0.0765)	1 (0.0658)	3 (0.0736)
0.07	2 (0.0799)	4 (0.0850)	1 (0.0755)	3 (0.0815)
0.08	2 (0.0894)	4 (0.0935)	1 (0.0840)	3 (0.0900)
0.09	2 (0.1039)	4 (0.1076)	1 (0.0982)	3 (0.1042)
0.10	2 (0.1316)	4 (0.1357)	1 (0.1265)	3 (0.1325)
0.11	2 (0.1615)	4 (0.1706)	1 (0.1593)	3 (0.1678)
0.12	2 (0.1829)	3 (0.1930)	1 (0.1813)	4 (0.1977)
0.13	2 (0.1971)	4 (0.2059)	1 (0.1939)	3 (0.1993)
0.14	2 (0.2030)	4 (0.2125)	1 (0.2015)	3 (0.2049)
0.15	2 (0.2027)	4 (0.2118)	1 (0.2012)	3 (0.2046)
0.16	2 (0.2125)	4 (0.2213)	1 (0.2106)	3 (0.2140)
0.17	1 (0.2147)	4 (0.2216)	2 (0.2150)	3 (0.2156)
0.18	2 (0.2197)	4 (0.2273)	1 (0.2191)	3 (0.2213)
0.19	2 (0.2355)	4 (0.2424)	1 (0.2336)	3 (0.2361)
0.20	3 (0.2521)	4 (0.2572)	1 (0.2503)	2 (0.2509)
Average	1.8571 (0.1317)	3.4762 (0.1369)	1.3810 (0.1301)	3.2857 (0.1363)

4.2. Experimental design

The SVM-HMM is a popular hybrid model which has been successfully used in speech recognition and human activity behaviour recognition. In the experiments, we compare our proposed MOGP-HMM with several SVM-HMMs. Specifically, we investigate SVMs using different kernels including radial basis function (RBF), polynomial and sigmoid kernels, denoted by SVM(R), SVM(P) and SVM(S), respectively. For further technical details about SVMs please refer to Cristianini and Shawe-Taylor (2000). The corresponding hybrid models are denoted by SVM(R)-HMM, SVM(P)-HMM and SVM(S)-HMM in the following discussion.

In the first stage, the synthetic data is used. We employ the 5-fold *cross-validation method* in training SVMs; and 50% of the data for training and 50% of data for validation in obtaining the MOGP algorithm (called the *test set* method). The training settings of the MOGP algorithm are summarised in Table 1. The experiments run up to 80,000 tree evaluations, each of which generates a new tree model. The training terminates when either of the following two conditions is met: (i) 0/1 loss converges; (ii) the maximum iteration number is achieved. The GP trees are initialised with ramped half-and-half method (Koza, 1992); point crossover and mutation are used. The tree depth is set as 4, and the tree node types include unary minus, addition, subtraction, multiplication, and analytic quotient (Ni, Drieberg, & Rockett, 2013). It should be noted that the test set method (Bishop, 2007) ensures the generalization capability of the MOGP. Regularization is difficult to implement for tree-based models as they are heuristic algorithms. In broader sense, regularization for tree-based models is proceeded by limiting the maximum depth of trees, ensembling more than just one tree, or setting stricter stopping criterion on when to split a node further (e.g. the minimum gain, the number of samples). Therefore, the above training steps and settings in Table 1 ensure the MOGP will not be over-fitting. In the second stage, the original real lifelogging data is firstly processed to obtain the corresponding observations. We then employ the 10-fold cross-validation method in training and testing the HMM, where the data is divided into 10 equal folds – 9 folds are used for estimating the parameters $\{\pi, \mathbf{A}, \mathbf{B}\}$ of the HMM and the remaining fold is used for prediction and evaluation. In the experiments, we use the HMMlib C++ in the implementation of the HMM (Sand, Pedersen, Mailund, & Brask, 2010).

4.3. Results and discussion

Fig. 6 presents an example of using the MOGP-HMM for a participant. The time series plot of the original real lifelogging data shows the values of steps, duration, and distance over time. The MOGP algorithm then classifies the collected lifelogging data into one of 6 classes (or states) representing the observations. The HMM then predicts the user's physical activity status over time based on the observations. The proposed MOGP-HMM is compared with other three SVM-HMMs for 10 participants under 21 noise levels, which gives 840 performance results in total. Table 2 presents the models' performance for one user. In machine learning theory, test error (also known as the generalisation error) is a measure of how accurately a model is able to predict outcome values on a set of data that it has never seen before. Test error and overfitting are considered to be closely related. Generally, the more overfitting occurs, the larger the test error. In each model, the test error increases with the increase of the noise. Under different noise settings, test errors of all four models are close. However, the SVM(P)-HMM leads the rankings, slightly ahead of the MOGP-HMM. Both models are significantly ahead of the other two models. Fig. 7 provides the results of overall performance for all 10 people in our data. The MOGP-HMM can achieve comparable performance as SVM-HMMs as it has the second smallest average test error in all four models.

We would like to provide some insights on the underlying differences between the MOGP algorithm and SVMs. From the perspective of model generalisation, test error in SVMs is proportional to the combination of training error and model complexity. Conducted from the structure risk minimisation scheme, a SVM converges to a linear optimal solution. As for the non-linear models, a kernel function is employed to non-linearly map the original feature space into a kernel space where the linear classifier is trained. Thus, the optimality of the linear model holds only in the kernel space that relies on the kernel function. Intuitively, kernels incur different non-linearity from one to another. Each optimal model from a specific kernel is an effective local optimum with respect to the kernel function used. Therefore, different results can be obtained with SVMs using different kernels. On the other hand, the MOGP algorithm minimises empirical 0/1 loss and the size of the tree simultaneously, leading the evolutionary process to minimise test error. Each GP tree represents a discriminant
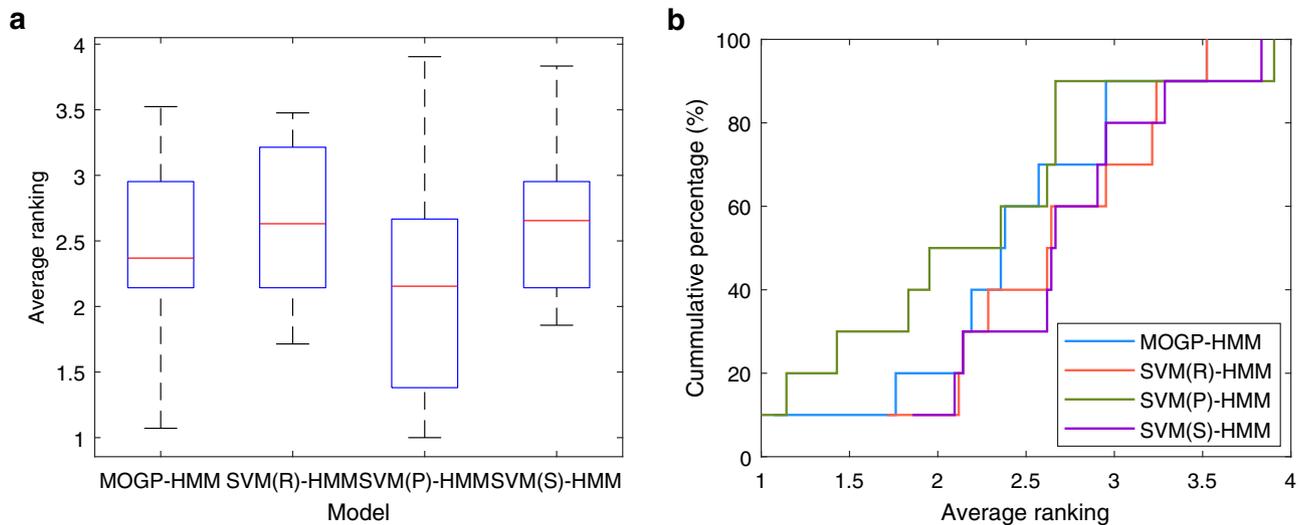
Fig. 7. Overall performance of hybrid models for all participants: (a) the boxplot of expected rankings; (b) the cumulative probability distribution of the average ranking.

that maps the training data from the feature space into a decision space while using a threshold to separate two classes. Compared to SVMs, one advantage of GP algorithms is that the discriminant is a syntax tree providing rich model candidates to search. The evolutionary process is driven by the MOGP algorithm towards a set of solutions non-dominant to each other in terms of empirical error and complexity. The solution set has no quantitative justification related to the expected risk. As a result the optimisation process is not as solid as SVM. Specifically, the tree size considered as a syntactic complexity measure is not as tightly coupled to the true complexity as Vapnik-Chervonenkis (VC) dimension employed in SVM (Cristianini & Shawe-Taylor, 2000). Therefore, the model evolved by MOGP is only a close-to-optimum result over a larger model space. Overall, the MOGP algorithm is a robust choice in the classification-HMM type hybrid models as no ad-hoc kernels are required and is underpinned by its flexible non-linearity.

5. Conclusion

In this paper, we propose a hybrid model MOGP-HMM to predict human physical activity status from sequential lifelogging data. The MOGP algorithm transforms the collected lifelogging data into observations, which are the input of the HMM. The latter is a chain-structured Bayesian network where the latent variables represent an individual's physical activity status over time. Given a sequence of observations, an individual's physical activity status can be predicted. We validate the proposed model with the real data collected from a group of participants in the UK, and compare our model with several SVM-HMMs in which SVMs use different kernels. Our experimental results show that the MOGP-HMM can achieve comparable performance as SVM-HMMs.

The contribution of our study is multi-fold. It contributes to the recent use of operational research, machine learning, data mining, big data and the Internet of things in healthcare. Lifelogging data collection and analysis is a big data problem and the developed model is a personalised data-driven model tailored to individual's physical activity pattern. We aim to achieve patient-centred healthcare where patient will play more active roles and be encouraged positive attitudes towards healthy lifestyles. As illustrated in Fig. 1, the proposed hybrid model can be used as a decision support tool that provides real-time health monitoring, statistical analysis and personalized advice to an individual through portable digital devices. Therefore, the study fits seamlessly with the current trend in the UK healthcare transformation of patient

empowerment as well as contributes to a strategic development for more efficient and cost-effective provision of healthcare. Using a MOGP algorithm in the two-stage hybrid structure has methodological contributions. It is non-parametric and can find an optimal trade-off between model fitness and complexity by setting the tree size. Unlike SVMs, it is not sensitive to the choice of kernel functions and thus provides more robust and discriminative representations of sparse data. To the best of our knowledge, this is the first study that uses a MOGP algorithm as a multi-class classifier to construct a classification-HMM hybrid model for solving sequential learning problems. Our model can be of interest and easily adapted to other relevant domains in business analytics such as consumer choice modelling and high dimensional business data classification or dimension reduction.

Acknowledgments

References

Abraham, A., Corchado, E., & Corchado, J. (2009). Hybrid learning machines. *Neurocomputing, 72*(13–15), 2729–2730.

Auria, L., & Moro, R. A. (2008). Support vector machines as a technique for solvency analysis. Deutsches Institut für Wirtschaftsforschung Technical Report.

Banaee, H., Ahmed, M., & Loutfi, A. (2013). Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors, 13*, 17472–17500.

Bishop, C. (2007). *Pattern recognition and machine learning*. Springer.

Blanchet, J., Gallego, G., & Goyal, V. (2016). A Markov chain approximation to choice modeling. *Operations Research, 64*(4), 886–905.

Borrajo, L., Baruque, B., Ccorchado, E., Bajo, J., & Corchado, J. (2011). Hybrid neural intelligent system to predict business failure in small-to-medium-size enterprises. *International Journal of Neural Systems, 21*(4), 277–296.

Brailsford, S., & Vissers, J. (2008). OR in health. *European Journal of Operational Research, 185*(3), 901–904.

Brailsford, S., & Vissers, J. (2011). OR in healthcare: A European perspective. *European Journal of Operational Research, 212*(2), 223–234.

Burkardt, J. (2014). *The truncated normal distribution*. Florida State University, Department of Scientific Computing Technical Notes.

Caspersen, C., Powell, K., & Christenson, G. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Reports, 100*(2), 126–131.

Chen, V., Kim, S., Oztekin, A., & Sundaramoorthi, D. (2018). Preface: data mining and analytics. *Annals of Operations Research, 263*(1–2), 1–3.

Chen, Y., Mabu, S., Shimada, K., & Hirasawa, K. (2009). A genetic network programming with learning approach for enhanced stock trading model. *Expert Systems with Applications, 36*(10), 12537–12546.

Choi, B., Pak, A., Choi, J., & Choi, E. (2007). Daily step goal of 10,000 steps: a literature review. *Clinical & Investigative Medicine, 30*(3), 146–151.

Concha, O., Yi, R., Xu, D., Moghaddam, Z., & Piccardi, M. (2011). HMM-MIO: an enhanced hidden Markov model for action recognition. Proceedings of the computer vision and pattern recognition workshops.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other Kernel-based learning methods.* Cambridge University Press.

Dag, A., Oztekin, A., Yucel, A., Bulur, S., & Megahed, F. (2017). Predicting heart transplantation outcomes through data analytics. *Decision Support Systems, 94*, 42–52.

Dag, A., Topuz, K., Oztekin, A., Bulur, S., & Megahed, F. (2016). A probabilistic data–driven framework for scoring the preoperative recipient-donor heart transplant survival. *Decision Support Systems, 86*, 1–12.

De Caigny, A., Coussement, K., & De Bock, K. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research, 269*(2), 760–772.

Debaere, S., Coussement, K., & De Ruyckc, T. (2018). Multi-label classification of member participation in online innovation communities. *European Journal of Operational Research, 270*(2), 761–774.

Denoyel, V., Alfandari, L., & Thiele, A. (2017). Optimizing healthcare network design under reference pricing and parameter uncertainty. *European Journal of Operational Research, 263*, 996–1006.

Domingos, P. (2015). *The master Algorithm: How the quest for the ultimate learning machine will remake our world.* Allen Lane.

Doumpos, M., & Zopounidis, C. (2016). Editorial to the special issue "business analytics". *Omega: The International Journal of Management Science, 59*, 1–3.

Evans, J. (2017). *Business analytics: Methods, models, and decisions* (2nd). Pearson.

Fries, B. (1976). Bibliography of operations research in health care systems. *Operations Research, 24*(5), 801–814.

Gandek, B., Ware, J., Aaronson, N., Apolone, G., Bjorner, J., Brazier, J., et al. (1998). Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. *Journal of Clinical Epidemiology, 51*(11), 1171–1178.

Ghaddar, B., & Naoum-Sawaya, J. (2018). High dimensional data classification and feature selection using support vector machines. *European Journal of Operational Research, 265*(3), 993–1104.

Ghaddar, B., Sakr, N., & Asiedu, Y. (2016). Spare parts stocking analysis using genetic programming. *European Journal of Operational Research, 252*, 136–144.

Ghahramani, Z. (2001). An introduction to hidden Markov models and Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence, 15*(1), 9–42.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT.

Goonatilake, S., & Khebbal, S. (1995). *Intelligent hybrid systems.* John Wiley & Sons.

Grünig, R., & Kühn, R. (2013). *Successful decision-making: A systematic approach to complex problems.* Springer.

Gurrin, C., Smeaton, A., & Doherty, A. (2014). Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval, 8*(1), 1–125.

Han, S., Zhang, M., Li, P., & Yao, J. (2014). SVM-HMM based human behavior recognition. In Proceedings of the international conference on human centered computing (pp. 93–103).

Harris, S., May, J., & Vargas, L. (2016). Predictive analytics model for healthcare planning and scheduling. *European Journal of Operational Research, 253*(1), 121–131.

Hejazi, T.-H., Badri, H., & Yang, K. (2019). A reliability-based approach for performance optimization of service industries: an application to healthcare systems. *European Journal of Operational Research, 273*(3), 1016–1025.

Hindle, A., Hindle, G., & Hindle, T. (2013). Geographical modelling of patient episode flows and hospital catchment populations: a case study in Northern Ireland. *Health Systems, 2*(1), 53–60.

Hindle, G., & Vidgen, R. (2018). Developing a business analytics methodology: a case study in the foodbank sector. *European Journal of Operational Research, 268*(3), 836–851.

Koza, J. (1992). *Genetic programming: On the programming of computers by means of natural selection.* MIT Press.

Krischer, J. (1980). An annotated bibliography of decision analytic approaches to health care. *Operations Research, 28*(1), 97–113.

Kulev, I., Pu, P., & Faltings, B. (2016). Discovering persuasion profiles using time series data. In Proceedings of the neural information processing systems time series workshop.

Kunc, M., Harper, P., & Katsikopoulos, K. (2018). A review of implementation of behavioural aspects in the application of OR in healthcare. *Journal of the Operational Research Society.* doi:10.1080/01605682.2018.1489355.

Li, Y., Vo, A., Randhawa, M., & Fick, G. (2017). Designing utilization-based spatial healthcare accessibility decision support systems: A case of a regional health plan. *Decision Support Systems, 99*, 51–63.

Liao, L., Fox, D., & Kautz, H. (2005). Location-based activity recognition using relational Markov networks. In *Proceedings of the 19th international joint conference on artificial intelligence* (pp. 773–778).

Luque, R., Casilari, E., Morón, M.-J., & Redondo, G. (2014). Comparison and characterization of android-based fall detection systems. *Sensors, 14*(10), 18543–18574.

Micucci, D., Mobilio, M., & Napoletano, P. (2017). UniMiB SHAR: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences, 7*(10).

Mohameda, A., & Nair, R. (2012). HMM/ANN hybrid model for continuous Malayalam speech recognition. *Procedia Engineering, 30*, 616–622.

Mortenson, M., Doherty, N., & Robinson, S. (2015). Operational research from Taylorism to Terabytes: a research agenda for the analytics age. *European Journal of Operational Research, 241*(3), 583–595.

Ni, J., Drieberg, R., & Rockett, P. (2013). The use of an analytic quotient operator in genetic programming. *IEEE Transactions on Evolutionary Computation, 17*(1), 146–152.

Ni, J., & Rockett, P. (2014). Training genetic programming classifiers by vicinal-risk minimization. *Genetic Programming and Evolvable Machines, 16*(1), 3–25.

Pate, R., Pratt, M., Blair, S., Haskell, W., Macera, C., Bouchard, C., et al. (1995). Physical activity and public health: a recommendation from the centers for disease control and prevention and the American college of sports medicine. *Journal of the American Medical Association, 273*(5), 402–407.

Peddabachigari, S., Abraham, A., Grosan, C., & Thomas, J. (2007). Modeling intrusion detection system using hybrid intelligent systems. *Journal of Network and Computer Applications, 30*(1), 114–132.

Poli, R., Langdon, W., & McPhee, N. (2008). A field guide to genetic programming. http://www.gp-field-guide.org.uk.

Roumani, Y., Roumani, Y., Nwankpa, J., & Tanniru, M. (2018). Classifying readmissions to a cardiac intensive care unit. *Annals of Operations Research, 263*(1–2), 429–451.

Rouyendegh, B., Oztekin, A., Ekong, J., & Dag, A. (2018). Measuring the efficiency of hospitals: a fully-ranking DEA-FAHP approach. *Annals of Operations Research, 263*(1–2), 1–18.

Royston, G. (1998). Shifting the balance of health care into the 21st century. *European Journal of Operational Research, 105*(2), 267–276.

Rudner, J., McDougall, C., Sailam, V., Smith, M., & Sacchetti, A. (2016). Interrogation of patient smartphone activity tracker to assist arrhythmia management. *Annals of Emergency Medicine, 68*(3), 292–294.

Sand, A., Pedersen, C., Mailund, T., & Brask, A. (2010). HMMlib: a C++ library for general hidden Markov models exploiting modern CPUs. In *Proceedings of the 2nd international workshop on high performance computational system biology* (pp. 126–134).

Sandıkci, B., Maillart, L., Schaefer, A., Alagoz, O., & Robert, M. (2008). Estimating the patient's price of privacy in liver transplantation. *Operations Research, 56*(6), 1393–1410.

Shao, L., Liu, L., & Li, X. (2014). Feature learning for image classification via multiobjective genetic programming. *IEEE Transactions on Neural Networks and Learning Systems, 25*(7), 1359–1371.

Stadermann, J., & Rigoll, G. (2004). A hybrid SVM/HMM acoustic modeling approach to automatic speech recognition. In Proceedings of the international conference on spoken language processing.

Tako, A., & Kotiadis, K. (2015). PartiSim: a multi-methodology framework to support facilitated simulation modelling in healthcare. *European Journal of Operational Research, 244*(2), 555–564.

Terwee, C., Gerding, M., Dekker, F., Prummel, M., & Wiersinga, W. (1998). Development of a disease specific quality of life questionnaire for patients with Graves' ophthalmopathy: the GO-QOL. *British Journal of Ophthalmology, 82*(7), 773–779.

Topuz, K., Uner, H., Oztekin, A., & Yildirim, M. (2018). Predicting pediatric clinic no-shows: a decision analytic framework using elastic net and Bayesian belief network. *Annals of Operations Research, 263*(1–2), 479–499.

UK National Health Service (2015). Physical activity guidelines for adults. http://www.nhs.uk/Livewell/fitness/Pages/physical-activity-guidelines-for-adults.aspx.

Vapnik, V. (2000). The nature of statistical learning theory. *Statistics for engineering and information science* (2nd). New York: Springer.

Vilarinho, T., Farshchian, B., Bajer, D., Dahl, O., Egge, I., Hegdal, S., Lønes, A., Slettevold, J., & Weggersen, S. (2015). A combined smartphone and smartwatch fall detection system. *IEEE International Conference on Computer and Information Technology.*

Wermter, S., & Sun, R. (2000). *Hybrid neural systems.* Springer.

Willis, G., Cave, S., & Kunc, M. (2018). Strategic workforce planning in healthcare: a multi-methodology approach. *European Journal of Operational Research, 267*, 250–263.

Won, K.-J., Hamelryck, T., Prügel-Bennett, A., & Krogh, A. (2007). An evolutionary method for learning HMM structure: prediction of protein secondary structure. *BMC Bioinformatics, 8*(1), 357.

World Health Organization (2014a). Global status report on noncommunicable diseases. http://www.who.int.

World Health Organization (2014b). Transforming health services delivery towards people-centred health systems. *Health Services Delivery Programme Briefing Note.*

World Health Organization (2017). Physical activity. *Health Topics.* http://www.who.int/topics/physical_activity/en/.

Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8*(2), 338–353.

Zhang, L., & Zhang, B. (2014). *Quotient space based problem solving: A theoretical foundation of granular computing.* Elsevier.

Zhang, M., Yin, P., Deng, Z., & Yang, D. (2008). SVM+BiHMM: A hybrid statistic model for metadata extraction. *Journal of Software, 19*(2), 358–368.

Zhang, Y., & Rockett, P. (2009). A generic multi-dimensional feature extraction method using multiobjective genetic programming. *Evolutionary Computation, 17*(1), 89–115.

Zhou, L., & Gurrin, C. (2012). A survey on life logging data capturing. *SenseCam Symposium.*

Zhou, Z. (2012). *Ensemble Methods: Foundations and Algorithms.* Chapman and Hall/CRC.