Presenting a Data Imputation Concept to Support the Continuous Assessment of Human Vital Data and Activities

Marian Haescher
Fraunhofer IGD
Rostock, Germany
marian.haescher@igd-r.fraunhofer.de

Denys J.C. Matthies
Augmented Human Lab,
The University of Auckland,
New Zealand
d.matthies@auckland.ac.nz

Silvio Krause
Fraunhofer IGD
Rostock, Germany
silvio.krause@vcric.igd-r.fraunhofer.de

Gerald Bieber
Fraunhofer IGD
Rostock, Germany
gerald.bieber@igd-r.fraunhofer.de

ABSTRACT

Data acquisition of mobile tracking devices often suffers from non-continuous data streams and even invalid input data. These issues are very well pronounced with current wearables tracking the user's activity and vital data. Typical reasons include the short battery life time as well as the fact that body-worn tracking device may be doffed. In addition, technical issues can arise that can corrupt the data and thus render it unusable. In this paper, we introduce a practical data imputation concept demonstrating how to fix incomplete datasets by using a new merging approach. We believe this approach to be particularly suitable for assessing activities and vital data. With our technique the dataset becomes coherent and comprehensive, so it is ready for further analysis. In contrast to previous approaches, our technique enables the controlled creation of continuous datasets that also contain information on the level of uncertainty for possible reconversions, approximations, or later analysis.

CCS CONCEPTS

• Human-centered computing → Ubiquitous and mobile computing design and evaluation methods; Ubiquitous and mobile computing systems and tools; Empirical studies in ubiquitous and mobile computing; • Design Methodology → Pattern Recognition;

KEYWORDS

Data Imputation; Data Fusion; Sensor Fusion; Controlled Data creation; Mobile Device; Smartwatch; Accelerometer; Coherent database.

ACM Reference format:

Marian Haescher, Denys J.C. Matthies, Silvio Krause and Gerald Bieber. 2019. Presenting a Data Imputation Concept to Support the Continuous Assessment of Human Vital Data and Activities. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '19). ACM, New York, NY, USA, 7 pages.* https://doi.org/10.1145/3316782.3322785

PETRA '19, June 5–7, 2019, Rhodes, Greece

ACM ISBN 978-1-4503-6232-0/19/06…$15.00
https://doi.org/10.1145/3316782.3322785

INTRODUCTION

Physical activity recognition and vital data analyses can be seen as enabler technologies to improve personal health [21]. Since current sensor-based systems can provide objective data about the user's current state, we are able to complement anamnesis information and improve self-estimations. In fact, these sensor systems have already permeated our everyday lives, which occur in any shape of smart consumer products such as smartphones, smartwatches, smartbands, etc. These new types of wearables incorporate a great variety of sensors, such as accelerometers, gyroscopes, PPGs, etc., and thus enable a convenient patient monitoring [20]. However, we face certain difficulties, e.g., when assessing gathered data that is incomplete or of poor quality. These issues can quickly occur with any type of body-worn sensor, such as pulse chest straps, which are usually just used for short periods of time (e.g., jogging or cycling). Other types of fitness trackers are also used intermittently and thus, the recording is not continuous. The most frequent down times of smartbands and smartwatches are caused by battery issues, since batteries still require charging periods of multiple hours.

Figure 1. Our data imputation concept enables a gapless data assessment of vital and activity data. We demonstrate our system on an Android smartwatch (No 1 D5+) as well as on a web server-based UI.

Visualizing, processing, and evaluating activity and vital data that are not continuous likely can lead to vague impressions and even false implications. For instance, an averaged daily activity level based on step counting cannot provide a good estimation when the device runs out of battery or is not worn at all. A similar problem occurs when looking at a graph displaying the sleep duration over the span of a week. However, when the device is not worn, one have to deal with incomplete datasets. Especially in elderly people, datasets demonstrate to be highly incoherent and fragmented, also because the device is forgotten to be worn. Analyzing such data is challenging. Reconstructing missing data is important to gain a realistic impression of the overall trend, and it supports decision making, e.g., in dosing medications, establishing best-fit fitness plans, or improving therapy. To resolve this issue, we require a general concept to merge available data to rectify erroneous data or to complement missing datasets. We propose a practical data imputation technique for activity and vital data using information from previous days. We present a new merging approach which transforms fragmented datasets into one coherent dataset (*Figure 2*). We argue that our technique leads to a realistic representation; however, we are aware that the imputation with generated data cannot reconstruct ground truth data. Focusing on health tacking, reconstructing data is especially critical, since the implications of how the data is used are more severe. One must be careful when concluding medical suggestions based on interpolated data. The suggested approach of data imputation introduced in this paper provides complete datasets that are more likely to be true and enable a gapless assessment, which is not possible with previous solutions that solely rely on recorded data. To ensure the dataset's accuracy, our concept indicates and distinguishes between ground truth data and interpolated data, which is introduced in this paper. This allows the application even in a medical context. Moreover, we outline the contribution of our paper by addressing the following research questions:

RQ1: Can we rationally merge fragmented datasets into a continuous dataset?

RQ2: What is a meaningful way of considering and weighing datasets with older time stamps?

RQ3: How can we distinguish between original and interpolated datasets?

RQ4: What are the benefits of the proposed merging concept for the analysis of activity and vital data?

RELATED WORK

The need for adequate methods that cope with missing data has a long history and is related to the current application field of the recognition of activity or vital data. There are three types of missing data: *Missing at Random (MAR)*, *Missing Completely at Random (MCAR)* and *Missing Not At Random (MNAR)* [1] – all types can occur in a health tracking scenario. It has been the aim of many works to gain complete datasets by applying different approaches. The terminology is not standardized, and thus, different notions, unclear definitions, and overlaps are described in the following:

Data imputation is the broader term for concepts that aim at generating almost realistic data and completing faulty datasets. The validity of the data depends on the algorithm used for the interpolation process. *Controlled data creation* describes the process of generating very similar datasets by using pattern recognition and machine learning algorithms. The term *data complementation* is used when incomplete datasets are filled with data which already exists somewhere else. Further, *data fusion* describes techniques that put heterogeneous types of data together to generate new data. This allows extracting new information that did not exist before. The term *data merging* is used when referring to different data sources in order to complete datasets. In this paper, we are using the broader term *data imputation* since we believe it to be the best fit.

Another term, *data curation*, is used for the organization, integration and maintenance of data collected from various sources. Data curation focuses on handling data in terms of data life span, reliability of data, and data elimination because of redundancy or data size issues. Data curation can be performed after data imputation is completed.

We now provide a rough overview of the state-of-the-art in the field of data generation, data imputation, and data curation, in reference to various application scenarios.

In the application field of *big data analysis*, merging datasets and eliminating redundant and outdated information is highly relevant. Therefore, a rule-based system is commonly used [2] to carry out data cleaning, as well as an incremental merge / purge algorithm. Even for big data, we may deal with continuously available sensor data. Here, a time dependence can be implemented to enhance the performance [3] of recurrent data, i.e. when data is redundant or repeats itself only with minor changes. Sometimes, similar data are generated by heterogeneous resources at the same time. Therefore, the data sources have to be integrated into one homogeneous dataset. The requirement of merging and combining data is also addressed in the field of *data integration*. Here, data from other databases are extracted, transformed and finally loaded into a new database (*Figure 2*). This concept of Extract, Transform, Load (ETL) is one of the general underlying concepts for the refreshment of the data warehouse content [4].

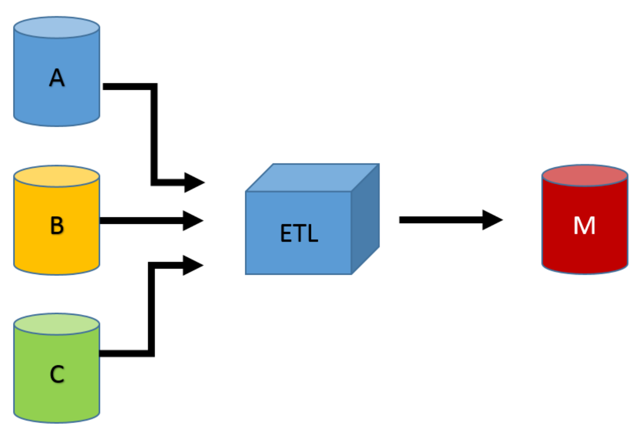

Figure 2. Extract, Transform, Load (ETL) Data Integration

The concept of ETL can be extended by semantic integration, since heterogeneous databases are interrelated. If databases consist of the same data type and data origin, only missing data are to be estimated. Here, *data curation* provides tools for *controlled data creation*, *data maintenance*, and *data manage* [5]. *Controlled data creation* can be accomplished by converting existing data into a new dataset [6].

In accordance to literature, the most common approach to estimate and reconstruct missing data is to make use of simple techniques, such as applying *Nearest-Neighbor*, *BayesNet*, *Gaussian Mixture* and *Decision Trees* [7]. More complex models include a *Multi-Layer Perceptron*, *Genetic Algorithms*, and *Particle Swarm Optimization* [8]. As it is difficult to know in beforehand which technique is suitable to impute data, the validity of the reconstructed data may be questionable.

The existing imputation techniques can be categorized by whether longitudinal data are used (in particular, periodic time series) or categorical data. Physical activity varies throughout the day, i.e., the physical activity depends on the time of day and the day of the week. For example, Tudor-Locke et al. [9] clustered the activities in weekdays and weekend days, though we believe further distinctions are required to reasonably structure the data. If recordings contain missing data, standard methods of data estimation include deterministic models, stochastic models, state space models, and curve fitting methods for missing values [10]. These standard models, however, are not suitable for this application field because although activity or vital data are assessed over a long period of time, the span of missing data can also be long, up to hours, days or even weeks. Thus, standard models may lead to wrong statistical or probabilistic parameters.

A possible imputation method for life span data was presented in 2016 by Halpin [11], using binary, multinomial or ordinal logistic regression models, and resulting in datasets that are free of missing data (even though a restitution of original data is hardly possible). Furthermore, the data imputation process in time series can be guided visually and statistically [12], and also, methods of artificial intelligence can be used for missing data imputation [13]. To reify the data imputation methods, it is useful to consider the special needs of different application fields.

Many behavioral studies have been using activity sensors. In particular, early research used electromechanical pedometers. On the one hand, this hardware had no operation time limitation. On the other hand, the database quality depended on the compliance of the users, e.g., on whether the hardware is being worn or being doffed. For monitoring physical activity, a doffed sensor system cannot distinguish between 'the user is active but no sensor is worn' and 'the user is inactive'. In those older studies, a self-estimation of the usage time and the average of recorded steps per day were used to identify dropout days and to raise the validity of the recorded data [14]. More recently, smarter systems automatically detect the state of being doffed, e.g., smartphones that are being used to support the treatment for obese patients [15].

Ideally, people wear their mobile monitoring devices almost permanently throughout the day. However, the average compliance rate is lower, for example, less than 90 % in a study with healthy subjects [10]. Even the compliance rates of elderly users are significantly below 100%, despite their high motivation to wear safety systems to ensure their own well-being and independence. Especially for medical studies, we require reliable and accurate data. Usually, for clinical studies, a larger set of data is assessed to allow for dropping out unusable datasets [17], inflating the overall effort. Fitness and lifestyle data, in contrast, are assessed sporadically and over a longer period. The analysis of the data should be as good as possible but this data contains no medical relevance.

In conclusion, the strategy of data creation and merging the created data into an existing dataset is dependent on the application field. In circumscribed application fields, a merging of datasets has been successfully performed in literature [18]. However, when it comes to assessing vital signs and activity, the enrichment of unmonitored episodes with controlled data creation to receive a continuous dataset is yet unavailable. Applying a data creation and merging concept for activity and vital data, we should rely on ideas proposed with previous methods, so data does not become invalid.

1. GENERAL CONCEPT

The general idea of our controlled data creation concept, enabling a continuous assessment of activity data and vital signs, is to complete the dataset by data being most likely to occur in that situation. The ideal concept would be identifying the user's intention for a situation, so we can make a good guess on what kind of activity or vital data is missing. In the application field of activity recognition, we assume the simplest user model: the weekends differ from the weekdays. Activities during the weekend are more similar to each other than in comparison to weekdays. Apparently, knowing the user's intention of action would enable the detection of recurrent activities, e.g., working out at the gym at a certain weekday. However, this requires an intelligent user model. If such user model does not exist, we can fall back on chronological occurrences. For this, we are breaking up the time-dependent data into segments, e.g., weeks, days, hours, etc. If datasets within these segments are missing, we look back to previous segments (e.g., at the same time from the previous days), and weighing the data (newer data are more relevant than older data).

3.1 Data Completion

In the field of activity and vital data, we are using the criterion "data continuity". Data continuity is the ratio of how many data is missing in relation to the regarded interval. If the dataset is complete, the continuity is 1; if no data was recorded in the time interval of interest, the continuity is 0. The proposed algorithm aims to optimize the criterion of fragmented datasets, generating a continuous dataset with no gaps / no missing data. For this, the algorithm is adding data that likely represents the missing data periods. The concept of data imputation intends that the outcome always provides a full set of continuous data. Missing data are replaced by known data of other time periods or situations. Figure 3 illustrates the concept of the data imputation if no recorded data is available at a specific time. The upper horizontal row describes activity data (blue and green) of a given time period, and the blank spaces are missing data. The middle row describes activity data of another time period, also with missing data (blank space). The algorithm merges the rows into a continuous dataset. The lowest row illustrates the result of the merging of the first two rows.

A possible simple implementation within time frames comparable in structure (e.g., during weekdays) is the transferring of yesterday's activity data into missing datasets of today. This method, for example, will enhance the representation of overall sleep duration of a workweek if a recording of one night failed. With this approach, the average daily sleeping hours and total time in bed are more exact as a simple summation of recorded data as it is done as of yet.

Figure 3. Activity diagram illustrating the merging of data collected over two days, with different activities in blue, green and yellow, and doffed states in white.

3.2 Time-Dependent Data Fusion

Supplementary to the data complementation using time stamps, the data imputation algorithm also uses the likelihood of the occurrence of the data (Figure 4). Regarding the strategy for data imputation and weighing, we consider three different scenarios, listed from strongest to weakest: absolute timing is similar, relative timing is similar, and order is similar. The general concept has to be specified for the specific aim of data analysis. We propose that (in case no user or behavior model exists) the basic weighting by time lag is sufficient.

Figure 4. Weighing activity data by time lag.

3.3 Parameter-Dependent Data Fusion

Activity and vital data do not only depend on time but also on the location and other pre- or post-conditions. These parameters support the identification of the user's intention. If we know the user's intention and behavior, we can create the most likely dataset to complement missing data. To optimize the data quality, we have to define and identify a goal criterion for the selection of the best strategy for controlled data creation.

3.4 Estimation of Uncertainty

After filling periods of missing data, it is useful to estimate the accuracy of the complemented datasets. For reasons of comparability, we define the uncertainty as 0 if every data point is valid and no insertion of additional data has occurred. Furthermore, we define the level of uncertainty as 1 if we fully approximate the data at random. The uncertainty is between 0 and 1 if we use data from other time stamps or if we use other datasets, e.g., sets of comparable users. Furthermore, the level of uncertainty relies on the data of interest. This leads to the interval of [0...1] for the level of uncertainty, with a rising value indicating rising uncertainty. The consideration is somehow weak because at this moment, we have no clear definition of uncertainty between zero and one. Still, the introduction of a level of uncertainty provides the advantage that a reverse data curation is possible at any time. The dataset can be easily transformed back to the originally recorded dataset with its gaps and missing periods.

For a representation of the level of uncertainty, a graph can be printed into an activity chart as follows (Figure 5):

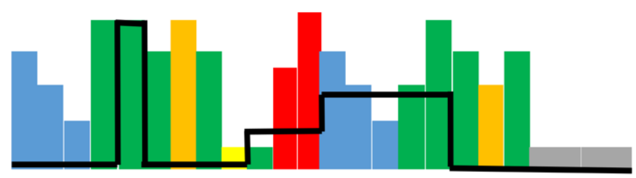

Figure 5. Level of uncertainty for activity and vital data (black graph).

2. IMPLEMENTATION

For evaluation purposes, we implemented a data recording application for smartwatches to assess human vital and activity data. We tracked and automatically classified different types of physical activity (resting, walking, running, and being active) as well as sleep and sleep intensity for every minute. In addition, we assessed pulse rate, respiration rate, sound level, and skin temperature every 10 minutes. The data was synchronized with a web server.

For each dataset, we added an additional database value that contained the level of uncertainty (LU). The LU was set to zero in case of existing data.

(1) If (data exist) then LU = 0;

If the data was missing for a very short time interval (arbitrarily set to 10 minutes), the last known value was used and received a level of uncertainty of 0.5. We were using 0.5 initially, because this is the exact middle of the interval of the level of uncertainty. In the course of our implementation and analysis, we identified this value as a suitable parameter.

Figure 6. Typical 24h-Activity diagram with sleep (blue), charging period (gap), activity period (green and orange) and doffed state (grey).

(2) If (matching predecessor <u>or</u> successor data from the same day) then LU = 0.5;

If the subsequent value also equals the last known value, the level of uncertainty is processed by:

(3) If (matching predecessor <u>and</u> successor data from the same day)
*then LU = 0.5 * 0.5;*

If data from 24h before (same timeslot) and the predecessor and successor data from the same day are matching, LU is computed as follows:

(4) If (matching data from the previous day in the same timeslot <u>and</u> matching predecessor <u>and</u> successor data from the same day)
*then LU = 0.5 * 0.5 * 0.5;*

If only data from 24h before or older are existing, their level of uncertainty is used for determining the current level of uncertainty. The uncertainty of previous days is progressing recursively (i.e., going stepwise backwards until data for the specific timeslot is found). To account for the uncertainty arising from the lack of data from the same day, a basal uncertainty of 0.5 is added:

(5) If (only old data)
*then LU = 0,5 + 0.5 * (LU of previous day or a previous day recursively progressing);*

If the old data was already uncertain, the level of uncertainty rises by this equation from day to day.

The algorithm stops with the inclusion of old data when recorded data has been found. A future work would be to compute the most likely activity out of the total of the available activities for this time interval.

Although this approach does not integrate transition probabilities, user behavior or overall activity limits, it provides continuous datasets of activity and vital data.

3. EVALUATION

To evidence that the improved datasets represents the user's behavior more likely, a preliminary evaluation is given as follows.

5.1 Goal

One goal of activity and vital data monitoring is to receive a general understanding about the user's fitness profile and to identify anomalies or prospects of improving the body condition. A fitness tracker provides many data but due to charging and other reasons, the data are mainly fragmented. A daily step graph may vary but not because of changes in walking behavior but because of the (dis)continuity of wearing the device and receiving data from the device. The aim of the evaluation was to receive a general understanding of the main reasons for fragmented data are and how often datasets are missing.

5.2 Method

For the evaluation, we identified one subject who was instructed to wear a smartwatch with fitness tracking capabilities for 3 months. That subject should wear the device constantly even during night. Since the subject should get accustomed to the watch, we chose the last month of wearing for the data analysis. We assume that the behavior with and handling of the smartwatch had become normal and stable during the last month. The hardware was a Smartwatch No.1 D5 with an Android operation system.

An errorless and continuous monitoring of activity and vital data with smartwatches is hard to achieve. While some special devices, such as actigraphs [19], provide a long battery life of two weeks, smartwatches have to be charged every 1-3 days, so the recording is bound to be fragmented.

5.3 Results

The evaluation period spans the described month, consisting of 30 days, which is a possible recording time of 43,200 minutes. The recorded valid data were ca. 32,539 minutes, i.e., 75.32 percent. This percentage may seem sufficient for an activity or sleep analysis, but only 14 fully complete days had been recorded. The other days were somehow fragmented and suffered from missing data. The time used for charging was automatically recorded by our application on the watch. We measured 647 minutes of charging time while the smartwatch was switched on. Some charging periods are missing because the application was not running when the device was switched off. So the charging time of the watch when it was switched off is unknown.

The smartwatch we used has a battery life of approx. 3 - 4 days with an energy-optimized recording application; the smartwatch was often in flight mode. The reasons of the gaps in the (supposedly continuous) data were as follows:

Technical reasons:

- *Charging time*
- *Technical failure, e.g., application crashes or lack of memory, operation system stops watch applications periodically*
- *Technical specifications, described below*

We noticed that the current operating system of the watch (Android 5.1, applied by No.1. D5+) sets the application on hold in a specified interval or restarts the application periodically. This may cause little gaps of data recording of approx. up to 2 minutes – unfortunately every 30 min. This behavior was not noticeable by the older hardware model of the same manufacture (No.1. D5).

Human Interaction:

- *No immediate wearing of the watch after recharging (the watch stays in docking station for additional time)*
- *After powering down (e.g., during the night), the user is not immediately recharging the watch*
- *The watch is doffed, e.g., because of washing dishes, taking a shower, or sport activities*
- *Forgetting to rebind the watch after having taken it off*

Other faulty human interactions, which have not occurred in this evaluation, may include accidently stopping the recording of the application, or forgetting a required manual input.

Figure 6 gives an example of the datasets assessed in the 24h-time frame, including periods of missing data. By using the proposed data imputation concept, the dataset was complemented, generating a continuous dataset.

Figure 7 illustrates the sleep probability over a week (7 days). The x-axis represents the time (0:00 a.m. - 12:00 p.m.), the y-axis describes how many times the subject was sleeping at the time of the day during the week. In the upper graph, we see a highly fragmented dataset, the lower graph represents merged data, which is smooth and continuous.

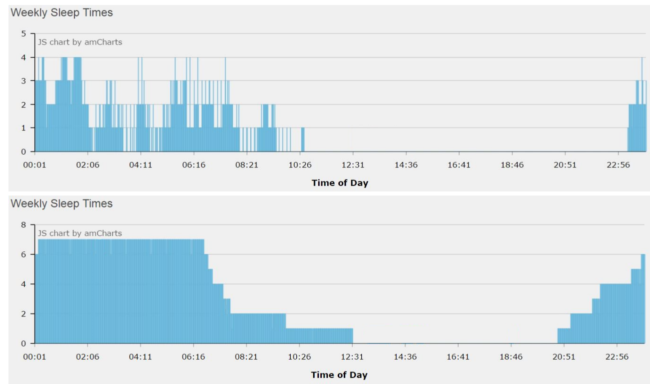

Figure 7. 24h-diagram of sleep probability without data merging (upper graph) and with data merging (lower graph)

5.4 Discussion

To our experience, the proposed concept of data imputation turned out to be a very useful technology to enrich datasets. Minor intervals of missing data due to discrepancies in operating system or software are handled very well. For instance, some data logging applications are recording in a specific time interval (e.g., 60 seconds) and this leads to the problem that sometimes, in one short interval, no data is recorded and in the short interval, two recordings occur. In our evaluation, the operating system was setting the recording sometimes on hold, which also resulted in missing data.

In many cases, a continuous dataset is needed to perform analytics or statistical interpretations. Some applications, though, may not be robust enough to handle the merging, which alters the data and changes the precision because estimated data are never true data. To meet this challenge, the inclusion of the proposed level of uncertainty is a very useful method to process the data. In our prototype application, we implemented two functions, one being the data imputation and the other being a de-imputation functionality, enabled by the uncertainty-level. Hereby, the added and estimated data will be erased again from the dataset and the original status is recovered. At every time, the user should be able to identify altered and true data visually, which is why we propose using another color or an additional graph for the level of uncertainty. The proposed method of data imputation provided a complete and sufficient dataset showing the overall activity level, sleep duration, time to bed and wake-up time distribution. We believe this data to be close to a realistic one, representing the activities similar to the actual one.

4. CONCLUSION

In this paper we presented a concept of recreating a continuous dataset out of fragmented, interrupted, and incoherent data. The proposed combination of merging existing data (preferably, from previous timestamps, or alternatively from similar users) and weighing them can be useful for both, activity data and vital data. We believe our concept to be also suitable for other types of datasets that are discontinuous and incomplete, e.g., diabetes datasets. We a straight-forward implementation of our algorithm. Moreover, we have shown that imputated data can be easily reorganized to restore the original datasets if required. We implemented and evaluated our proposed algorithm on an activity- and sleep-tracking service platform called "Diatrace", which only showed low computational extra load to the server. We see great potential in the adoption and optimization of other specific application fields. However, further research is required. Currently, we are working on the imputation of data that is only partly related to the user but not from the user's origin. We anticipate that data from other users with a similar user profile can be used to fill up open gaps. This may further enhance the results, which opens further research questions for follow-up research.

ACKNOWLEDGMENTS

This research has been supported by the German Federal State of Mecklenburg-Western Pomerania.

REFERENCES

[1] Rubin, D. B. (1976). Inference and missing data. In *Biometrika*, 63(3), 581-592., DOI: 10.1002/j.2333-8504.1975.tb01053.x

[2] Hernández, M. A., & Stolfo, S. J. (1998). Real-world data is dirty: Data cleansing and the merge/purge problem. In *Data mining and knowledge discovery,* 2(1), 9-37.

[3] Krishnan, N. C., & Cook, D. J. (2014). Activity recognition on streaming sensor data. In *Pervasive and mobile computing,* 10, 138-154.

[4] Vassiliadis, P. (2009). A survey of Extract–transform–Load technology. In *International Journal of Data Warehousing and Mining (IJDWM)*, *5*(3), 1-27.

[5] Stonebraker, M., Bruckner, D., Ilyas, I. F., et.al. (2013). Data Curation at Scale: The Data Tamer System. In *CIDR - Conference on Innovative Data Systems Research*, U.S.A., 2013.

[6] Coyle, K. (2012). Linked data tools: connecting on the Web. *American Library Association.*

[7] Yuan, Y. C. (2010). Multiple imputation for missing data: Concepts and new development (Version 9.0). *SAS Institute Inc, Rockville*, MD, 49, 1-11.

[8] Marwala, T. (Ed.). (2009). Computational Intelligence for Missing Data Imputation, Estimation, and Management: Knowledge Optimization Techniques: Knowledge Optimization Techniques. In *IGI Global.*

[9] Tudor-Locke, C., Burkett, L., Reis, J. P., Ainsworth, B. E., Macera, C. A., & Wilson, D. K. (2005). How many days of pedometer monitoring predict weekly physical activity in adults?. In *Preventive medicine*, 40(3), 293-298.

[10] Fung, D. S. (2006). Methods for the estimation of missing values in time series, *Dissertation, Edith Cowan University, Australia.*

[11] Halpin, B. (2016). Multiple imputation for categorical time-series, *The Stata Journal 16(3),* pp. 590-612

[12] Bögl, M., Filzmoser, P., Gschwandtner, T., Miksch, S., Aigner, W., Rind, A., & Lammarsch, T. (2015, October). Visually and statistically guided imputation of missing values in univariate seasonal time series. In *IEEE Conference on Visual Analytics Science and Technology (VAST 2015)*, pp. 189-190. IEEE.

[13] Marwala, T. (2009). Computational Intelligence for Missing Data Imputation, Estimation, and Management: Knowledge Optimization Techniques: Knowledge Optimization Techniques. In *IGI Global,* Chicago.

[14] Tudor-Locke, C., Williams, J. E., Reis, J. P., & Pluto, D. (2002). Utility of pedometers for assessing physical activity. In *Sports Medicine*, 32(12), 795-808.

[15] Schiel, R., Kaps, A., & Bieber, G. (2012). Electronic health technology for the assessment of physical activity and eating habits in children and adolescents with overweight and obesity IDA. In *Appetite*, 58(2), 432-437.

[16] Mathie, M. J., Coster, A. C., Lovell, N. H., Celler, B. G., Lord, S. R., & Tiedemann, A. (2004). A pilot study of long-term monitoring of human movements in the home using accelerometry. In *Journal of telemedicine and telecare,* 10(3), 144-151.

[17] Acebo, C., Sadeh, A., Seifer, R., Tzischinsky, O., Hafer, A., & Carskadon, M. A. (2005). Sleep/wake patterns derived from activity monitoring and maternal report for healthy 1-to 5-year-old children. In *SLEEP-New York then Westchester,* 28(12), 1568.

[18] Schröder, M., Bader, S., Krüger, F., & Kirste, T. (2016). Reconstruction of everyday life behaviour based on noisy sensor data. In *Proceedings of the 8th International Conference on Agents and Artificial Intelligence*, 430-437. SCITEPRESS-Science and Technology Publications, Lda.

[19] Actigraph, 49 East Chase Street. Pensacola, FL 32502, U.S.A., www.actigraphcorp.com.

[20] Haescher, M., Matthies, D.J.C., Srinivasan, K., and Bieber, G. (2018). Mobile Assisted Living: Smartwatch-based Fall Risk Assessment for Elderly People. In *Proceedings of the 5th International Workshop on Sensor-based Activity Recognition and Interaction (iWOAR '18'),* Berlin, Germany. ACM. https://doi.org/10.1145/3266157.3266210

[21] Trimpop, J., Haescher, M., Bieber, G., Matthies, D.J.C., Lämmel, F., & Burggraf, P. (2015). The Digital Health Companion: Personalized Health Support on Smartwatches via Recognition of Activity-and Vital-Data. In *Proceedings of International SUMMER SCHOOL on Visual Computing 2015 (SCVC '15),* Rostock, Germany (p. 125). Fraunhofer Verlag. https://doi.org/10.13140/RG.2.1.3500.8083