

Summer 6-15-2016

CAPTURE AND ANALYSIS OF SENSOR DATA FOR ASTHMA PATIENTS

Nikolai Bock

University of Applied Sciences Mainz, nikolai.bock@hs-mainz.de

Matthias Scholz

University of Applied Sciences Mainz, matthias.scholz@hs-mainz.de

Gunther Piller

University of Applied Sciences Mainz, gunther.piller@hs-mainz.de

Klaus Böhm

health&media GmbH, Darmstadt, klaus.bohm@health-media.de

Follow this and additional works at: http://aisel.aisnet.org/ecis2016_rip

Recommended Citation

Bock, Nikolai; Scholz, Matthias; Piller, Gunther; and Böhm, Klaus, "CAPTURE AND ANALYSIS OF SENSOR DATA FOR ASTHMA PATIENTS" (2016). *Research-in-Progress Papers*. 2.
http://aisel.aisnet.org/ecis2016_rip/2

This material is brought to you by the ECIS 2016 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in Research-in-Progress Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

CAPTURE AND ANALYSIS OF SENSOR DATA FOR ASTHMA PATIENTS

Research in Progress

Bock, Nikolai, University of Applied Sciences Mainz, Mainz, Germany,
nikolai.bock@hs-mainz.de

Scholz, Matthias, University of Applied Sciences Mainz, Mainz, Germany,
matthias.scholz@hs-mainz.de

Piller, Gunther, University of Applied Sciences Mainz, Mainz, Germany,
gunther.piller@hs-mainz.de

Böhm, Klaus, health&media GmbH, Darmstadt, Germany, klaus.bohm@health-media.de

Abstract

Worldwide more than 230 million people suffer from asthma. Reliable and timely guidance for individuals to minimize their risk for asthma attacks is not available. This is largely due to the fact that asthma symptoms are often caused by multiple environmental and personal factors. Many of them are neither captured nor systematically analysed. This is addressed by the project ActOnAir. It aims at a comprehensive capture of health factors and the environmental exposure of individuals, as well as a subsequent analysis in real-time. For this purpose the ActOnAir system provides a mobile sensor box for data collection, a sensor data integration and processing platform, a data mining component and a smartphone application for patients. This contribution outlines the design objectives of the ActOnAir system and discusses corresponding key requirements. The related system architecture is introduced and first results from a prototype implementation are sketched.

Keywords: mobile eHealth systems, asthma, sensing and monitoring, people as sensors, real-time analytics

1 Introduction and Motivation

Air pollution is a major health risk. Estimates of the World Health Organization reveal that approximately one in eight of total global deaths result from the exposure to air pollutants. In particular people with asthma – approximately 235 million – are affected by worsening air quality (WHO, 2015). Although there is plenty of research on the influence of environmental factors upon the appearance of asthma symptoms (see e.g. Schachter et al., 2015), detailed personal guidance for everyday life is hard to provide. Reasons are:

- A comprehensive and fine granular capture of bio-signals of individuals and their exposure to environmental factors is missing. Many pollutants are only measured in a few places and difficult to relate to individual persons.
- Individual analyses and situational predictions are absent. The dependence of asthma attacks on pollution thresholds, environmental aspects and individual dispositions are not considered in a holistic way.

In the near future available information on personal bio-signals and the environment will strongly increase due to the growing number of mobile and stationary sensors. Therefore, it is all the more important to develop methods and tools for a reliable provisioning of situation specific, individual guidance.

The situation described above shall be improved significantly through a new hard and software system “ActOnAir”. A corresponding project has been started early 2015. This contribution describes the project objective and first results. In focus are the system architecture and a proof of concept implementation of the software system. The approach follows the design science research method (Hevner et al., 2004).

2 State of the Art

Since around 10 years local and fine granular measurements of air pollutants are topics of a growing number of projects. Examples are CitiSense (2010) and Copenhagen Wheel (2014). Both measure the concentration of different environmental factors through mobile sensor systems, covering for example ozone, nitrogen oxides, carbon monoxide, temperature and relative humidity. Furthermore, new concepts and trends, like People as Sensors (see e.g. Sagl et al., 2015) and Quantified Self, where individuals collect data about their life using technological tools (see e.g. Quantified Self, 2016), yield more and more information on personal wellbeing. A shortcoming of many past analyses is that they do not combine environmental data with information on personal wellbeing (for a recent discussion see e.g. Resch and Blaschke, 2015). In addition, a flexible extension of sensor systems with further data sources has not been considered in the projects mentioned above. The ActOnAir system addresses both issues: Environmental and personal data are integrated in a meaningful way, despite the inherent heterogeneity of the corresponding data sources. Additionally, extendable integration services allow the subsequent integration of new sensor data.

Data mining of sensor information becomes increasingly important for personalized analyses in health care. The development of methods which cover all potentially relevant data and ensure an appropriate interpretation and further processing of results is high on the agenda (see e.g. Sow et al., 2013). A good starting point for asthma monitoring is the contribution of Lee et al. (2010). In this work data mining methods for the joint analysis of personal bio-signals and environmental factors were investigated. The considered approaches integrate sequential pattern mining for the extraction of features of asthma attacks with the identification of classifiers through decision tree mining or a rule-based method, respectively. Sequential pattern mining has been investigated and applied to other areas in healthcare too. For example in Reps et al. (2012) general patient data have been analysed to highlight patients susceptibility to future illness based on their medical history and personal data. In Wright et al. (2015) predictions for next prescribed medications have been calculated. In the above studies data mining algorithms have been applied off-line to data collected and stored in the past. Real-time data mining of continuously generated data streams from patient devices has been studied in Abadia et al. (2011). In ActOnAir off-line sequential pattern mining and classification is combined with a real-time application of decision trees. This approach allows real-time forecasts based on clearly interpretable rules. These can systematically be refined to continuously improve the guidance of individual patients.

3 Objectives and Requirements

The final design artifact of our research project is a hard and software system which overcomes the current deficits described above. The corresponding design objectives are:

- O1: The burden of individuals suffering from asthma will be captured in a comprehensive and detailed way. The personal exposure of patients to air pollutants and environmental factors will be measured in short time intervals and correlated to up-to-date individual health factors.

- O2: Personal guidance to patients shall be provided in real-time, tailored to the constitution and the current situation of the individuals.

These objectives are starting point for the identification of requirements for the IT system – the design artefact. This contribution is focusing on the architecture of the corresponding software system. The emphasis is on the interplay of its core components with a direct relation to the design objectives. Further aspects are mentioned only, if required for an overall understanding. Requirements on data security and privacy are not part of this discussion.

Input for the ActOnAir software system are continuously measured data for different air pollutants and environmental factors, as well as information about the personal health of patients. For their capture a mobile unit (mobile sensor box) with existing and newly developed sensors will be provided. Details about this hardware component will be described elsewhere. In addition, data from the area of Quantified Self will be used. Further information about personal wellbeing will be collected through a dedicated mobile application – applying the concept People as Sensors. Besides data from individuals, generally available information from weather and air quality stations will be integrated too.

In addition to the mobile sensor box, the overall IT system is made up of the following logical building blocks: A component to integrate and manage sensor data (SD), a building block for the analysis of measured data with various data mining methods (DM), and a mobile application for patients (MA). In the following the main requirements for these components are introduced, starting with sensor data management:

- SD1: Integration of all sensor data, including technical and semantical cleansing and harmonization. Combination of sensor data measured by individual persons and data from generally available information sources
- SD2: Integration of data from new sensors or services without significant development effort, i.e. plug-and-play
- SD3: Real-time provisioning of actual data to mobile applications MA for the execution of forecasts
- SD4: Continuous provisioning of harmonized sensor data to the data mining component DM for analysis

Requirements SD1 and SD2 relate to the design objective O1, while SD3 is a prerequisite of O2. SD1, SD2 and SD4 are essential for the derivation of forecast rules within the data mining component DM.

DM itself needs to serve to the following key requirements:

- DM1: Retrieval of actual and historic environmental and health data from the SD component. Classification of data into different person segments. Execution of data mining analyses within defined segments
- DM2: Communication of individual forecast models from data mining for real-time guidance to the mobile applications MA of end users
- DM3: Identification of optimization potential for personal forecast models based on the evaluation of data obtained from the sensor data management component SD

The uniform processing of all sensor data in the data mining analysis supports design objective O1. DM2 and DM3 aim at O2. A growing data basis may, for example, lead to more fine granular person segments and an eventual complete individualization of forecasts. DM2 serves also real-time forecasts which are part of O2.

Building upon the results of the data mining component, the mobile application MA must be able to:

- MA1: Provisioning of real-time guidance for end users based on forecasting models from the data mining component and actual sensor data
- MA2: Individual guidance for end users without sharing personal data in cloud environments
- MA3: High usability of personal forecasts, including the possibility of a personalization of results to serve individual preferences

Requirement MA1 focusses towards design objective O1. MA2 and MA3 are most important for the general acceptance of the software system by end users.

4 System Architecture

A suitable system architecture is shown in Figure 1. It consists of five main building blocks:

Mobile Sensor Box: To measure the individual exposure of persons to air pollutants, the ActOnAir system provides a wearable box with a set of sensors. Amongst others it contains measuring units for particulate matter emissions PM10 and PM2.5, for ozone, temperature and relative humidity. Captured data are transmitted via Bluetooth to the mobile applications of end users and from there to the module Sensor Integration.

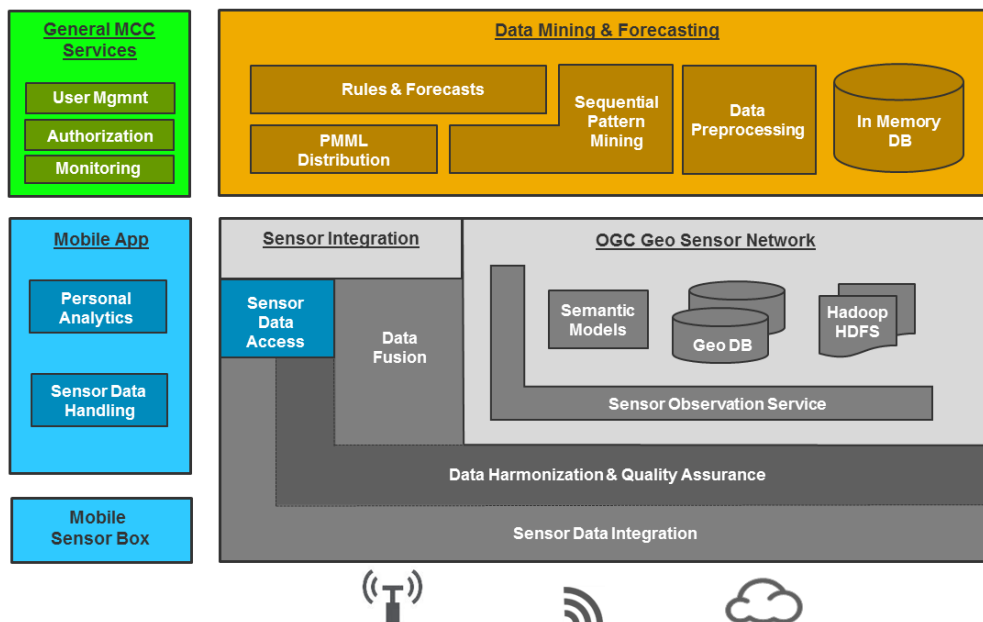


Figure 1: ActOnAir system architecture

Sensor Integration and Geo Sensor Network: This component is responsible for the processing of heterogeneous sensor data. The module Sensor Integration provides services for a flexible plug-and-play integration and combination of data from different sources. Harmonized sensor data are persisted within the Geo Sensor Network. It is based on standards of the Sensor Web Enablement from the Open Geospatial Consortium (OGC), like the Sensor Observation Services (SOS) and SensorML (OGC, 2015). The component Data Fusion enables the combination of sensor data, personal bio-signals and generally available environmental information. For this purpose all data are projected onto appropriate spatial-temporal references.

Data Mining and Forecasting: This building block receives harmonized and interpolated sensor data from the Data Fusion component. During pre-processing, data are assigned to suitable patient seg-

ments. Data mining analyses then yield frequent sequential patterns for days with asthma attacks, as well as for times without discomfort. These patterns serve as input for the derivation of rules for forecasts. Corresponding results are finally provisioned to end user applications through a document formatted according to the Predictive Model Markup Language (PMML). The system shall be able to automatically identify potential improvements for individual forecasts, e.g. based on newly measured sensor data. It may then appropriately optimize patient segments, improve the binning of data, calculate new rules and distribute them accordingly.

Mobile Application: The user or patient interacts through a smartphone application with the ActOnAir system. The mobile app calculates and displays personal forecasts. For this purpose it interprets the rules from the data mining component using actual sensor data and personal bio-signals. In addition, the smartphone application enables a regular recording of personal symptoms through an asthma diary. It typically covers information on peak expiratory flow, cough, wheeze, shortness of breath, chest tightness and further indications. Finally, the smartphone application acts as control unit for all mobile and personal sensors.

Mobile Cloud Computing Services (MCC): This module provides common mobile cloud computing services for, e.g., user management, authorization as well as communication and application monitoring.

The interplay of these major components of the ActOnAir system is illustrated through a typical data flow: The smartphone app of a patient collects information on her asthma symptoms as well as sensor data captured by her Mobile Sensor Box. All data are transferred to the component System Integration. They are then technically and semantically harmonized. Finally, they are persisted in the Geo Sensor Network. Here also publicly available environmental data, e.g. from weather and air quality monitoring stations, are stored. In the module Data Fusion environmental data are extrapolated to appropriate temporal and spatial coordinates and combined with data from individuals. As a result, one obtains comprehensive datasets with a best fit to the space-time trajectory of patients.

Next, the datasets are provided to Data Mining and Forecasting. After appropriate pre-processing, e.g. segmentation and binning, frequent sequential patterns and rules for forecasting are derived. The rules for forecasting are transferred to the smartphone application of a patient. Here they are used for real-time guidance. Inputs are: actual symptoms, most current sensor data from the Mobile Sensor Box, possible health data from other wearable devices of a patient, and general environmental data with proper temporal and spatial reference. The latter are retrieved on request from interfaces of the component Sensor Data Access.

In which way the illustrated architecture serves the described key requirements is illustrated through two examples:

- **Integration of new sensors:** To integrate the data from new sensors (requirement SD2), the messaging patterns of Hohpe and Woolf (2003) have been implemented. They enable the transfer of sensor data through standardized methods from different channels to pre-defined endpoints for further processing. Examples for such endpoints are sensor adapters or data fusion processes for sensor data enrichment. The implementation of these messaging patterns also supports the flexible orchestration of process chains for, e.g., harmonization, filtering, fusion or storage of sensor data. As a consequence, the effort to integrate new sensors or measurement concepts is limited to the implementation of corresponding adapters for new data structures and a potential extension of processing workflows, e.g., through NLP algorithms. This satisfies the requirement for flexible sensor integration and extension (Bock and Boehm, 2015).
- **Real-time guidance and model optimization:** To enable real-time guidance of individuals based on current environmental data and bio-signals (requirement DM2), the analysis methods within the data mining and forecasting component have been split: With sequential pattern mining highly frequent sequences of measured data values for air pollution, weather conditions and individual health factors before days with and without asthma attacks are identified in an asynchronous way. These

frequent patterns are then used as features for the identification of decision trees (Lee et al., 2010). After their transfer to mobile end-user applications, they can be used to forecast individual risks for asthma attacks in real-time, based on most current sensor data.

This separation and distribution of analysis components also allows a continuous improvement of forecasting models (requirement DM3): The quality of models can be regularly inspected in an asynchronous way. If, e.g., the constantly growing data volume of measured sensor data leads to significant model improvements for particular patient segments, new models shall be transferred to the corresponding end user applications for better real-time forecasting.

5 Prototype Implementation

This section summarizes selected details of a prototype implementation of the ActOnAir software system. For the Geo Sensor Network the SOS framework of 52°north (52n SOS, 2015) is used. It implements the actual OGC SOS standards and appropriate information models for air quality. The SOS framework allows a physical as well as a virtual integration of data from different sources. Based on current OGC standards (OGC, 2015), the system offers a uniform API for outgoing and incoming data communication, covering also spatiotemporal filter operations. The component Sensor Integration uses SpringXD (Spring XD, 2015), which builds upon Spring Integration and the messaging patterns of Hohpe and Woolf (2004). Within a first proof of concept adapters for generally available services for environmental data were implemented. They typically use different formats, like CSV, JSON or NetCDF. All basic services for data processing and communication with mobile applications and the component for data mining have been established and tested.

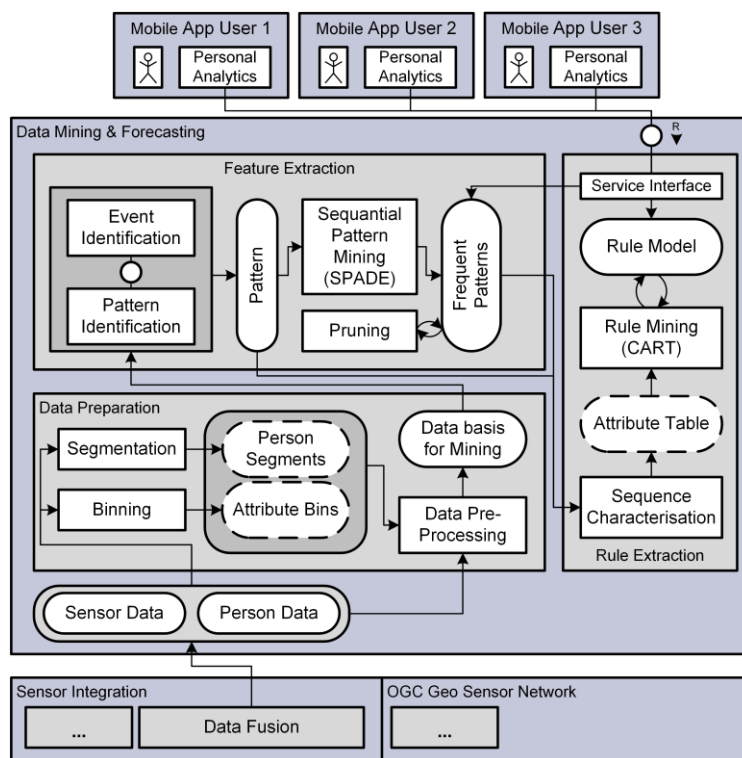


Figure 2: Architecture of the component Data Mining and Forecasting

The module Data Mining and Forecasting is built upon the in-memory platform SAP HANA. The performance of SAP HANA is beneficial for the expected high data volumes: If 5% of asthma patients in

Germany would use the application and would collect sensor data four times an hour, the approximate data volume sums up to about 10 GB per day, or several TB per year. To obtain short response times for initial explorative analyses as well as systematic improvements of forecasting models, the programming model is chosen appropriately: A de-normalized data model has been selected; table variables are used as far as possible, to avoid write operations; data intensive application logic has been largely embedded into the database with SQLScript; stored procedures – e.g. for data pre-processing – have been parallelized wherever applicable. In Figure 2 examples for virtual tables are indicated for person segments, attribute bins and attribute tables. Also various areas for parallelization can be pointed out: The functional components for binning, segmentation, data preprocessing, pattern identification and sequential pattern mining can be executed independently in parallel for different sensor types and individual patients.

Also the Predictive Analysis Library (SAP PAL, 2015) within the HANA platform is of significant advantage, as ActOnAir needs to use and combine several data mining algorithms in a performant way. For sequential pattern mining an R implementation of the SPADE algorithm is applied (Buchta et al., 2015). Frequent patterns for sensor data, environmental information and personal symptoms are taken as attributes for the computation of decision trees (Lee et al., 2010). For this purpose the CART algorithm within PAL has been used (SAP PAL, 2015). The corresponding results are finally provided to the smartphone applications of end users as PMML documents through an OData interface.

This multi-step analysis procedure has been successfully tested with sample data for selected asthma symptoms and generally available weather data and environmental information.

6 Next Steps

In the area of sensor data management algorithms for the harmonization and fusion of diverse sensor data will be implemented next. Also necessary is the completion of the mobile application for end users. Furthermore, an authorization concept based on OAuth will be explored for the overall data flow.

For data mining and forecasting test runs with realistic data are scheduled. For the fine-tuning of the applied algorithms, a medical evaluation of results is of particular importance. Potential adjustments include the segmentation of patients, binning of data during pre-processing, pruning of frequent patterns and the simplification of decision trees.

In addition, concepts for an automated optimization of personal predictions will be investigated. Results from a retrospect comparison of predictions with corresponding event data from patients, or significant changes in volume and quality of sensor data could be important triggers.

In Q3 2016 the complete system should be available for final evaluation.

Supported by the German Federal Ministry for Economic Affairs and Energy and the HPI Future SOC Lab.

References

- Abadia, R., Stranieri, A., Quinn, A., Seifollahi, S. (2011). “Real-time processing of data from patient biodevices.” In: *Proceedings of the Fourth Australasian Workshop on Health Informatics and Knowledge Management*. Ed. by K. Butler-Henderson and T. Sahama. CRPIT Volume 120, p. 25.
- Bock, N. and Boehm, K. (2015). “Flexible Datenintegration und Migration in ein Sensornetzwerk - ein Beispiel mit Twitter-Analysedaten.“ *AGIT - Journal für Angewandte Geoinformatik* 1, 448-453.
- Buchta, C., Hahsler, M., Diaz, D. (2015). *Package ‘arulesSequences’*. URL: <https://cran.r-project.org/web/packages/arulesSequences/arulesSequences.pdf/> (visited on 11/16/2015).

- CitiSense (2010). *CitiSense*. URL: <http://sosa.ucsd.edu/confluence/display/CitiSensePublic/CitiSense/> (visited on 11/16/2015).
- CopenhagenWheel (2014). *The Copenhagen Wheel*. URL: <http://senseable.mit.edu/copenhagenwheel/> (visited on 11/16/2015).
- Hevner, A. R., March, S. T., Park, J., Ram, S. (2004). "Design science in information systems research." *MIS Quarterly* 28 (1), 75-105.
- Hohpe, G., Woolf, B. (2003). *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Boston: Addison-Wesley.
- Lee, C. H., Chen, J. C., Tseng, V. S. (2010). "A Novel Data Mining Mechanism Considering Bio-Signal and Environmental Data with Applications on Asthma Monitoring." *Computer Methods and Programs in Biomedicine* 101 (1), 44-61.
- OGC (2015). *OGC Standards and Supporting Documents*. URL: <http://www.opengeospatial.org/standards/> (visited on 11/16/2015).
- Quantified Self (2016) *Quantified Self*. URL: <http://quantifiedself.com/> (visited on 03/07/2016).
- Resch, B. and Blaschke, T. (2015). "Fusing Human and Technical Sensor Data: Concepts and Challenges." *SIGSPATIAL* 7, 29-35.
- Sagl, G., Resch, B., Blaschke, T. (2015). "Contextual Sensing: Integrating Contextual Information with Human and Technical Geo Sensor Information for Smart Cities." *Sensors* 15, 17013-17035.
- SAP PAL (2015). *SAP HANA Predictive Analysis Library*. URL: http://help.sap.com/hana/sap_hana_predictive_analysis_library_pal_en.pdf/ (visited on 11/16/2015).
- Schachter, E. N, Moshier, E., Habre, R., Rohr, A., Godbold, J., Nath, A., Grunin, A., Coull, B., Koutrakis, P., Kattan, M. (2015). "Outdoor air pollution and health effects in urban children with moderate to severe asthma." *Air Quality, Atmosphere & Health*, 1-13.
- Sow, D., Turaga, D. S., Schmidt, M. (2013). "Mining of sensor data in health care: a survey." In: *Managing and Mining Sensor Data*. Ed. by C. C. Aggrawal. Heidelberg: Springer, p. 459.
- Spring XD (2015). *Spring XD*. URL: [http://projects.spring.io/spring-xd.](http://projects.spring.io/spring-xd/) (visited on 11/16/2015).
- WHO (2013). *Chronic respiratory diseases, Asthma*. URL: <http://www.who.int/respiratory/asthma/en/> (visited on 11/16/2015).
- Wright, A. P., Wright, A.T., McCoy, A., Sittig, D. (2015). "The use of sequential pattern mining to predict next prescribed medications." *Journal of Biomedical Informatics* 53, 73-80.
- 52n SOS (2015). *Sensor Web Community*. URL: <http://52north.org/communities/sensorweb/index.html/> (visited on 11/16/2015).