Sensing, Processing and Analytics – Augmenting the UBICON Platform for Anticipatory Ubiquitous Computing

Martin Atzmueller
University of Kassel
ITeG Research Center Kassel,
Germany
atzmueller@cs.uni-kassel.de

Björn Fries
University of Kassel
ITeG Research Center Kassel,
Germany
fries@cs.uni-kassel.de

Naveed Hayat
University of Kassel
ITeG Research Center Kassel,
Germany
nha@cs.uni-kassel.de

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).
Ubicomp/ISWC'16 Adjunct, September 12–16, 2016, Heidelberg, Germany
ACM 978-1-4503-4462-3/16/09.
http://dx.doi.org/10.1145/2968219.2968438

Abstract
Anticipatory systems require different steps like sensing, data processing, context inference, and context prediction. Then, suitable platforms can support the implementation of the respective steps. This paper proposes an anticipatory ubiquitous perspective on the UBICON platform, considering data capture (sensing), localization (context inference) and activity recognition (context prediction), enabled by an integration of different technologies and tools. In an integrated approach, we propose different components for augmenting the UBICON platform. For these, we present results of respective case studies in ubiquitous and social environments. Our results demonstrate the applicability of the UBICON platform for these tasks, towards an extended platform for anticipatory ubiquitous computing.

Author Keywords
sensing; data processing; data analytics; ubiquitous social computing; network analysis; localization; human behavior; anticipatory computing; software platform

ACM Classification Keywords
H.1.m [Information Systems]: Models and Principles; H.2.8 [Information Systems]: Database Applications

Introduction

The emergence of social media and ubiquitous computing has created a number of novel social and ubiquitous environments. The data generated by these can then be analyzed, e. g., using data mining techniques. This also includes interaction data from sensors and mobile devices, as long as the data is created by real users.

With more and more of those ubiquitous devices emerging in our daily lives, sensor data capturing human activities is thus becoming a universal data source for the analysis of human behavioral patterns, and for building according models. However, as a prerequisite appropriate data capture and data processing methods need to be developed and combined into an integrated approach. Then, this combination enables ubiquitous and anticipatory computing integrating different but complementary methods, techniques and tools. Using these workflows from sensing to providing actions, e. g., for recommandations and proactive guidance can then be implemented.

This paper presents an extended view on the open-source UBICON platform [2][1], augmenting it with components for data capture, processing and analytics, as visualized in Figure 1; here different open-source tools are integrated like the sensor data collection framework (SDCF)[2] [5], the data analytics platform VIKAMINE[3] [8], the data mining toolkit WEKA[4] [25], and the GNU R environment[5] [21] for statistical computing: We sketch three case studies in the context of the the augmented UBICON platform. First, we present a method for resource-aware smart sensing, i. e., data capture using single-board architectures, as a flexible method

[1]http://www.ubicon.eu
[2]http://www.sdcf.eu
[3]http://www.vikamine.org
[4]http://www.cs.waikato.ac.nz/ml/weka/
[5]http://www.r-project.org

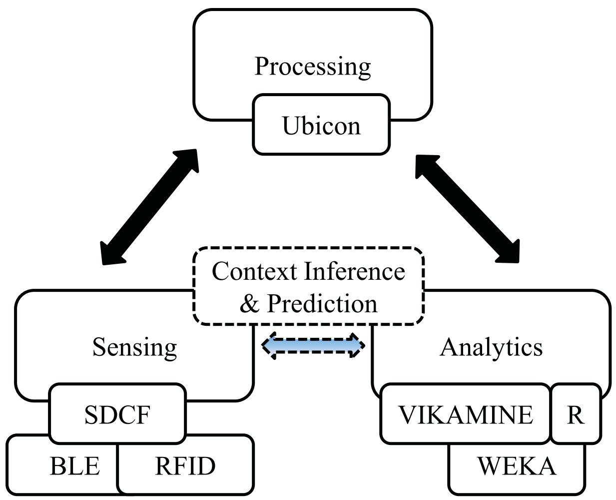

Figure 1: Augmenting UBICON for anticipatory ubiquitous computing: Integrated Sensing, Processing & Analytics.

to collect RFID data. Second, we propose an approach for indoor localization [24, 23], i. e., context inference utilizing Bluetooth low-energy (BLE) technology. We present results of an evaluation in a real-world setting and discuss implications. After that, we focus on context prediction using activity recognition techniques. Here, we propose a method for deriving interpretable rules for context prediction in order to enhance the understandability of the applied model.

The rest of the paper is structured as follows: We first provide an overview on the UBICON platform. After that, we present our case studies in context, i. e., summarizing the approaches and presenting evaluation results. Finally, we conclude with a summary and provide interesting further directions and options for future work.

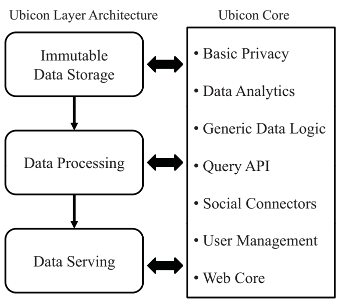

Figure 2: Conceptual overview on the architecture of the UBICON software platform, cf. [2].

UBICON

The UBICON software platform [2] aims at enhancing ubiquitous and social networking, for flexibly implementing applications in that context. It aims at supporting applications at the intersection of ubiquitous and social computing, integrating functionalities of both environments, providing efficient and effective approaches for building applications in areas like ubiquitous and social computing, internet of things, participatory sensing, and social crowd sourcing.

UBICON provides a number of components for data collection, processing, and serving. Grounded by fundamental principles of big data storage, processing, and analytics [19], UBICON features flexible ways for adaptions and extensions in the respective applications: At its core it can be extended, for example, by specialized components, e. g., for sensing and data analytics. Furthermore, it provides the means for creating and hosting customized applications.

System Architecture

Following [2], Figure 2 shows a conceptual overview on the architecture of the UBICON software platform. From a data-centric view, UBICON implements a data storage, processing, and serving pipeline similar to the *lambda architecture* [19] for handling and managing big data. In that way, core concepts such as immutability and recomputation are transparently enabled by the platform. Accordingly, the data flow is organized in the layers *immutable data storage*, *data processing*, and *data serving* providing flexible and transparent access to the data, e. g., for implementing big data analytics using Map/Reduce [12]. The functionality for each of these layers is backed by the UBICON core which provides canned functionality, i. e., framework classes and interfaces, which can be utilized throughout different applications. For a more detailed discussion of the architecture and its core concepts, we refer to Atzmueller et al. [2].

Overall, UBICON enables the observation of physical and social activities. Typically, applications utilize the provided core components, interfaces, and classes and extend the overall workflow according to their individual application requirements, e. g., as for the web applications developed in the EveryAware project[6], built on top of UBICON. Other applications include, for example, the Conferator [3], a social conference guidance system, and MyGroup, an application for enhancing social interaction in working groups. Both use active SocioPattern[7] RFID tags, which allow to localize participants and to collect their face-to-face contacts.

By including according data analytics components, the collected data can then be applied for anticipatory applications. Using collective sensing data, for example, obtained using the Conferator system, Atzmueller et al. [4] analyze the interactions and dynamics of the behavior of participants at conferences; similarly, the connection between research interests, roles and academic jobs of conference attendees is further analyzed in Macek et al. [18] which can then be applied for enhancing interactions at conferences, e. g., by recommendation systems [2]. In addition, we can provide indications about the currently *active* context, e. g., by visualizing anomalies concerning the current sensing information. Furthermore, connecting collective sensing data to online data, Scholz et al. [22] analyzed the predictability of links in face-to-face contact networks and additional factors also including online networks, which can then be applied for recommendation, and personalization. Further aspects in anticipatory systems include explanations. UBICON has been designed according to principles of explanation-aware software [9], and supports explanation capabilities of different components utilizing the platform. For more details on the latter, we refer, e. g., to [2].

[6] http://www.everyaware.eu
[7] http://sociopatterns.org/

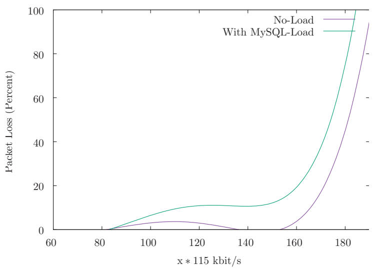

Figure 3: Performance/UDP-Loss with No-Load and MySQL load for a typical conference application (200 INSERTs/sec, see below).

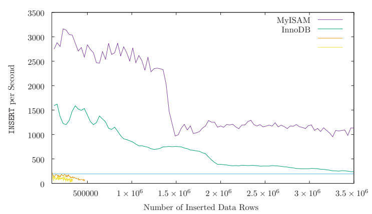

Figure 4: Scalability with increasing number of inserted rows using the iibench-mysql tool.

Case Studies

In the following, we briefly sketch three case studies on the augmented UBICON platform. First, we summarize a resource-aware component for sensing, i. e., RFID data capture using single board computers. Second, we discuss localization and present an evaluation specifically applying sensors that are available on mobile phones (i. e., BLE) as an alternative to RFID tags. Finally, we propose an activity recognition method for context prediction based on interpretable class association rules. The combination of these techniques enables ubiquitous anticipatory computing using UBICON, e. g., in conference contexts or in medical applications, where context of actors, their contacts, and individual behavior needs to be observed.

Resource-Aware RFID Data Capture

Sensing and data capture in RFID-based scenarios, i. e., using active RFID tags typically requires centralized server and storage paradigms. This implies, for example, that RFID readers that capture the data sent by RFID tags need to be connected to the server, e. g., by an Ethernet connection. This has significant impact on the application areas of the technology, e. g., due to environmental or building constraints where the respective system needs to be set up. For the UBICON platform, we experienced several situations for which the technical setup was difficult, e. g., when readers could not be installed in certain locations since Ethernet was not available etc. In addition, such static scenarios inhibit highly dynamic environments, for example, in which RFID readers capturing the signals can also be mobile.

We applied RFID tags developed by the SocioPatterns collaboration – providing an RFID infrastructure that detects close-range and face-to-face proximity (1-1.5 meters) of individuals wearing proximity tags with a temporal resolution of 20 seconds [10]. In contrast to, e. g., bluetooth-based methods that allow the analysis based on co-locationd data [5], here face-to-face proximity can be observed with a probability of over 99% using the interval of 20 seconds for a minimal contact duration. This infrastructure has been deployed in various environments so far, e. g., conferences [10, 4, 18], or workplaces [2].

A proximity tag sends out two types of radio packets: Proximity and tracking signals. Proximity radio packets are emitted at very low power and their exchange between two devices is used as a proxy for the close-range proximity of the individuals wearing them. Packet exchange is only possible when the devices are in close enough contact to each other (1-1.5 meters). The human body acts as an radio frequency shield at the carrier frequency used for communication [10]. The tracking signals are typically received by antennas of RFID readers installed at fixed positions in the facility environment. These tracking signals are then used to relay proximity information to a central server, using UDP.

As an alternative, we developed a component for augmenting UBICON for RFID data capture, specifically relying on a distributed infrastructure based on single board computers (SBC), cf. [14]. Then, the RFID readers that are necessary for capturing signals sent by the applied active RFID tags can operate stand-alone. A connection to a central server is not necessary any more, since the data is collected by the individual SBC and can be aggregated later. We performed several experiments concerning both performance and applicability of the technology, using a Raspberry PI2 for implementation. Considering performance we simulated data capture at a conference (see [18] for more details), and also used generated data for a benchmark comparison. In particular, we assessed possible UDP loss and the possible data throughput rate writing to a database backend (MySQL) at a typical conferencing scenario.

Overall, our results demonstrate the benefit and applicability of the proposed approach: Concerning UDP-loss we encountered a critical situation only for a data rate above 9MB/s which corresponds to 200 RFID tags sending to 20 RFID readers at the same time connected to the same SBC. This situation is unrealistic in our application scenarios, where we typically utilize fewer tags and also much fewer readers per SBC. In addition, we observed very good scalability with respect to the insert performance on the database using the iibench-mysql tool[8]: It inserts synthetic datasets into a table with two secondary indices and reports the respective performance of the last interval after 2500 inserts each. In that experiment, we also did not encounter the critical range of 200 INSERT/sec at all (which we empirically determined at a typical conferencing scenario). Also, utilizing real data collected at a conference was well in the supported performance range.

Localization

The localization of persons or objects in indoor environments is still a complex problem, especially with limited resources. Typically, the techniques are highly accurate in positioning, but expensive in hardware acquisition. In our setting, we use resource-aware and cost-effective active RFID tags, that are able to detect the proximity of other RFID tags. Using this technology is challenging, because only limited information is available for the localization task. We address this problem in UBICON [24, 23], where we developed new supervised and unsupervised algorithms that use proximity information for increasing the localization accuracy at room-level basis. Using proximity information, we could improve the localization accuracy from 84 % using a baseline algorithm to nearly 90 %, as evaluated during the poster session at the LWA 2010 conference. The algorithms are implemented in Java and GNU R.

Table 1: Accuracy of the localization experiments.

Algorithm	Accuracy
K-nearest Neighbor (k=5)	0.62
Logistic Regression	0.69
Naive Bayes	0.71
Random Forest	0.76

As an alternative, e. g., when RFID proximity tags are not available or cannot be deployed, we can utilize BLE modules. In a prototypical project, cf. [13], we evaluated an approach utilizing BLE [15], since BLE modules are typically available on modern smartphones. Using the AltBeacon library[9] we utilized BLE-USB sticks for implementing BLE beacons, transmitting beacon packets consisting of a UUID, two user-defined values, and information about the used transmission power at regular intervals (e. g., one packet per second).

In a first experiment, we applied four BLE-beacons for localizations of 10 areas (localization points) in a smart home context. We collected a dataset consisting of 200 measurements for each localization point, by measuring the signal strengths of the beacons at the different localizations for 200 seconds, each. This resulted in a dataset of 2000 measurements, for a fingerprinting approach. We evaluated different supervised machine learning methods, as implemented in the WEKA library for classification. In our experiments we used 10-fold cross validation, and estimated the accuracy of each method. The results are shown in Table 1.

Overall, we observe that the random forest classifier works significantly better than the other alternative, i. e., compared to k-nearest neighbors, logistic regression and naive bayes. Overall, the fingerprinting approach worked very well. However, we also considered locations that are relatively close: Here, for example, the k-nearest neighbor algorithm yielded many misclassifications, whereas the random forest classifier still provided robust results outperforming the competing approaches clearly. In our application context, the ensemble method given by random forest (using 10 decision trees) could adapt to the localization scenario best, also handling potential noise due to relatively close locations.

[8] https://github.com/tmcallaghan/iibench-mysql

[9] http://altbeacon.org

Class Association Rule Mining using Subgroup Discovery
In *activity recognition* approaches, often the learned models are either "black-box" models such as neural networks, or are rather complex, e. g., in the case of random forests or large decision trees. In this context, we propose to construct rule-based models for activity recognition. In particular, we focus on class association rules in order to aid interpretation by humans. Then, anticipatory models can be inspected, which is important, e. g., in medical or industrial applications.

Class association rules are special association rules with a fixed class attribute in the rule consequent. In order to mine such rules, we apply subgroup discovery [16, 26, 1] – an exploratory approach for discovering interesting subgroups defined by a *description*, e. g., a conjunction of attribute–value pairs (i. e., a typical rule body) with respect to a binary target concept. In the case of class association rules, the respective class can be defined as the target concept (i. e., the rule head). Then, subgroup discovery can be adapted as a rule generator for class association rule mining, using our CARMA framework presented in [6].

In summary, we compared an instantiation of the CARMA framework against two baselines: The Ripper algorithm [11] as a rule-based learner, and the C4.5 algorithm [20] for learning decision trees. We opted for interpretable patterns with a maximal length of 7 conditions. In the evaluation, we used three different *TopK* values: 100, 200 and 500. For the rule combination strategy, we experimented with four strategies: taking the best rule according to confidence and Laplace value, the unweighted voting strategy, and the weighted voting (Laplace) method, cf. [6] for a detailed discussion. All experiments were performed using 10-fold cross-validation on an activity dataset with 27 activities (classes) and 116 features, cf. [6] for details.

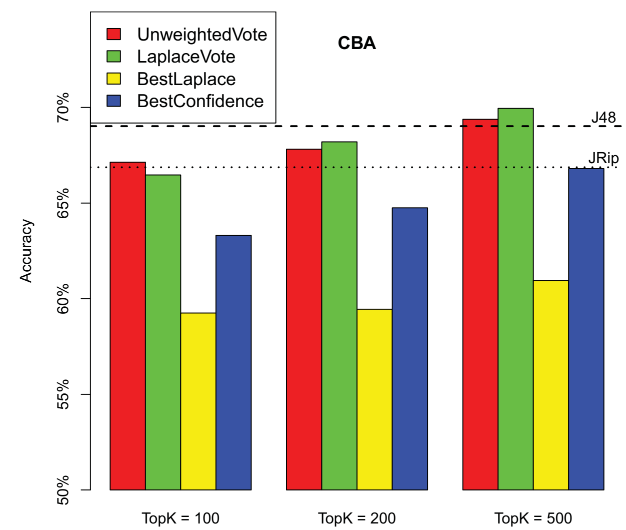

Figure 5: Comparison of the accuracy of CARMA using the standard CBA method for rule selection, with different rule combination strategies to the baselines.

Figure 5 shows the accuracy of CARMA using these parametrizations. Overall, it is easy to see that the proposed approach is able to outperform the baselines in accuracy. Furthermore, it outperformed both as well in complexity, since it always had a significantly lower average complexity regarding the average number of conditions in a rule. C4.5 showed a better performance than Ripper, however, with a more complex model (1394 rules) that were also more complex themselves; Ripper had a slightly lower accuracy but a signicantly lower number of rules and average rule length.

The proposed framework always provides a more compact model than the baseline algorithms: It is at least in the same range or even better. Considering the best parameter instantiation, the proposed approach is able to outperform both baselines concerning the accuracy and always provides a more compact model concerning rule complexity, cf. [6] for more details. Overall, this approach provides for a flexible activity recognition and context prediction method, e. g., in the scope of the SDC framework [5].

Conclusions

This paper provided an anticipatory ubiquitous computing perspective on the open-source UBICON platform, augmenting it with components for data capture, processing and analytics, demonstrated in three case studies: We presented a method for resource-aware smart sensing, proposed an approach for indoor localization [24, 23], and summarized an approach for context (activity) prediction using class association rules. These components form the basis for implementing various applications, e. g., in medical assistance contexts, where sensing (and localization) often needs to be implemented in resource-aware processing environments (enabled by UBICON), with explanation-aware analytics methods [9]. Here, the combination and tight integration of data capture, processing and analytics components in UBICON provides a flexible approach for implementations.

For future work, we aim to apply the augmented UBICON platform in more diverse scenarios, and to develop further analytics approaches, also by integrating heterogeneous data, e. g., [7, 17], in order to advance methods for intelligent feedback and proactive guidance for the users.

REFERENCES

1. Martin Atzmueller. 2015. Subgroup Discovery – Advanced Review. *WIREs DMKD* 5, 1 (2015), 35–49.

2. Martin Atzmueller, Martin Becker, Mark Kibanov, Christoph Scholz, Stephan Doerfel, Andreas Hotho, Bjoern-Elmar Macek, Folke Mitzlaff, Juergen Mueller, and Gerd Stumme. 2014. Ubicon and its Applications for Ubiquitous Social Computing. *New Review of Hypermedia and Multimedia* 20, 1 (2014), 53–77.

3. Martin Atzmueller, Dominik Benz, Stephan Doerfel, Andreas Hotho, Robert Jäschke, Bjoern Elmar Macek, Folke Mitzlaff, Christoph Scholz, and Gerd Stumme. 2011. Enhancing Social Interactions at Conferences. *it - Information Technology* 53, 3 (2011), 101–107.

4. Martin Atzmueller, Stephan Doerfel, Andreas Hotho, Folke Mitzlaff, and Gerd Stumme. 2012. Face-to-Face Contacts at a Conference: Dynamics of Communities and Roles. In *Modeling and Mining Ubiquitous Social Media*. LNAI, Vol. 7472. Springer, Berlin, Germany.

5. Martin Atzmueller and Katy Hilgenberg. 2013. Towards Capturing Social Interactions with SDCF: An Extensible Framework for Mobile Sensing and Ubiquitous Data Collection. In *Proc. MSM 2013, Hypertext 2013*. ACM Press, New York, NY, USA.

6. Martin Atzmueller, Mark Kibanov, Naveed Hayat, Matthias Trojahn, and Dennis Kroll. 2015. Adaptive Class Association Rule Mining for Human Activity Recognition. In *Proc. MUSE 2015*. Porto, Portugal.

7. Martin Atzmueller, Peter Kluegl, and Frank Puppe. 2008. Rule-Based Information Extraction for Structured Data Acquisition using TextMarker. In *Proc. LWA*. University of Wuerzburg.

8. Martin Atzmueller and Florian Lemmerich. 2012. VIKAMINE - Open-Source Subgroup Discovery, Pattern Mining, and Analytics. In *Proc. ECML/PKDD*. Springer, Berlin, Germany.

9. Martin Atzmueller and Thomas Roth-Berghofer. 2010. The Mining and Analysis Continuum of Explaining Uncovered. In *Proc. 30th SGAI International Conference on Artificial Intelligence (AI-2010)*.

10. Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. 2010. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS ONE* 5, 7 (07 2010), e11596.

11. William W. Cohen. 1995. Fast Effective Rule Induction. In *Twelfth International Conference on Machine Learning*. Morgan Kaufmann, 115–123.

12. Jeffrey Dean and Sanjay Ghemawat. 2008. MapReduce: Simplified Data Processing on Large Clusters. *Commun. ACM* 51, 1 (Jan. 2008), 107–113.

13. Björn Fries. 2015a. Localization using Bluetooth Low Energy. Project Report, Chair of Knowledge and Data Engineering, University of Kassel. (2015).

14. Björn Fries. 2015b. RFID-Infrastructure based on Single Board Computers: Evaluation of Performance and Applicability. Bachelor thesis, Chair of Knowledge and Data Engineering, University of Kassel. (2015).

15. Carles Gomez, Joaquim Oller, and Josep Paradells. 2012. Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-Power Wireless Technology. *Sensors* 12, 9 (2012), 11734–11753.

16. Willi Klösgen. 1996. Explora: A Multipattern and Multistrategy Discovery Assistant. In *Advances in Knowledge Discovery and Data Mining*. AAAI Press, 249–271.

17. Peter Kluegl, Martin Atzmueller, and Frank Puppe. 2009. Meta-Level Information Extraction. In *Proc. KI*. Springer, Berlin. (233–240).

18. Bjoern-Elmar Macek, Christoph Scholz, Martin Atzmueller, and Gerd Stumme. 2012. Anatomy of a Conference. In *Proc. ACM Hypertext*. ACM Press, New York, NY, USA, 245–254.

19. Nathan Marz and James Warren. 2013. *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*. Manning Publishers.

20. Ross Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.

21. R Development Core Team. 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

22. Christoph Scholz, Martin Atzmueller, Alain Barrat, Ciro Cattuto, and Gerd Stumme. 2013. New Insights and Methods For Predicting Face-To-Face Contacts. In *Proc. ICWSM*. AAAI Press, Palo Alto, CA, USA.

23. Christoph Scholz, Martin Atzmueller, and Gerd Stumme. 2014. Unsupervised and Hybrid Approaches for On-Line RFID Localization with Mixed Context Knowledge. In *Proc. ISMIS*. Springer, Berlin, Germany.

24. Christoph Scholz, Stephan Doerfel, Martin Atzmueller, Andreas Hotho, and Gerd Stumme. 2011. Resource-Aware On-Line RFID Localization Using Proximity Data. In *Proc. ECML/PKDD*. Springer, Berlin, Germany, 129–144.

25. Ian H. Witten and Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques* (2nd edition ed.). Morgan Kaufmann.

26. Stefan Wrobel. 1997. An Algorithm for Multi-Relational Discovery of Subgroups. In *Proc. 1st Europ. Symp. Principles of Data Mining and Knowledge Discovery*. Springer, Berlin, Germany, 78–87.