Poster Abstract: Sensor-based Clustering for Indoor Applications

Matthias Gauger, Olga Saukh, Marcus Handte, Pedro José Marrón
Universität Bonn, Bonn, Germany
Fraunhofer IAIS, St. Augustin, Germany
{gauger, saukh, handte, pjmarron}@cs.uni-bonn.de

Andreas Heydlauff, Kurt Rothermel
Universität Stuttgart, Germany
{heydlaas, rothermel}@ipvs.uni-stuttgart.de

Abstract—The clustering of nodes in wireless sensor networks is an important and widely used technique. We propose a clustering of nodes that reflects real world semantics meaningful to applications, i.e., based on room boundaries in indoor scenarios. We demonstrate the feasibility of automatically creating clusters that reflect rooms by analyzing sensor measurements with the help of statistical data clustering methods.

I. INTRODUCTION

Clustering is an important and widely used technique in wireless sensor networks. Common clustering techniques form groups mainly based on network criteria like, for example, connectivity information. However, a variety of applications can benefit from clusters of nodes that were created based on real-world criteria like room boundaries. An important example is the management of redundancy in the network, e.g., temporarily deactivating nodes to prolong the network lifetime. This can be done more safely and more effectively if it is known which nodes reside together in the same area. Other examples include role assignment as described in Frank and Römer [1], anomaly detection and room level querying.

We propose a novel method for the clustering of nodes in indoor scenarios that groups together nodes located in the same room. Our approach is based on analyzing the measurements of inexpensive, broadly available sensors with the help of statistical data clustering methods [2]. The motivation for this stems from the observation that sensor readings of nodes located in the same area often behave similarly and the expectation that these similarities can express themselves in a detectable correlation of the sensor values.

II. SENSOR-BASED CLUSTERING

A. Analysis

We perform a clustering of nodes based on sensor data as part of a four step process as illustrated in Fig. 1: After the inital **collection of sensor data** from all nodes, we obtain a clustering of nodes in three steps: **data preprocessing**, **similarity calculation** and finally **data clustering**. In the preprocessing of data and the calculation of the similarity, knowledge about the application domain can play an important role, e.g., when knowledge about the expected behavior of sensor values in certain time periods exists.
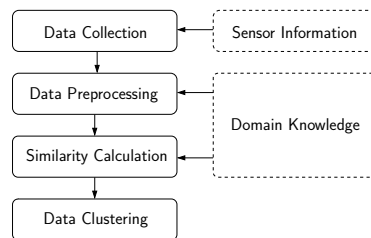

Fig. 1.   Processing steps of sensor-based clustering

Data preprocessing or data filtering modifies or transforms the acquired sensor data in a way that supports the following steps. It allows to compensate some effects of the lack of calibration, to emphasize specific features hidden in the data or to incorporate domain knowledge. We experimented with several methods including normalization, data smoothing and the detection of sensor events.

The next step, the *similarity calculation* determines the pairwise strength of the relationship among sensor nodes using the preprocessed sensor data vectors. Examples of methods for this calculation include the Euclidean distance, the Phi coefficient or Pearson's product-moment correlation coefficient.

The final step, the *data clustering* step, creates clusters of nodes based on the pairwise similarity of the nodes. While some methods calculate clustering trees (hierarchical clustering), others directly assign nodes to clusters (partitional clustering). In the analysis of data clustering we concentrated on approaches that can be calculated efficiently for the typical types of data generated in sensor networks.

At the beginning of our analysis, we collected time-stamped raw sensor data from various types of sensors in several scenarios over extended periods of time to check our assumptions concerning the correlations among sensor nodes and to investigate the applicability of different possible methods.

The analysis of the collected sensor data with different combinations of preprocessing, similarity calculation and data clustering methods confirmed our assumption that a room-level clustering of nodes based on simple sensor data is possible

with low overhead for different criteria. The analysis also allowed us to identify a set of particularly well-performing criteria that should be supported in a sensor-based clustering application.

B. Approach

Based on the promising results of the centralized analysis we designed a sensor-based clustering application. Important goals in its design included avoiding a centralized collection of sensor data samples, not requiring any (time) synchronization and minimizing the required message exchange among the nodes. We also aimed at minimizing the overhead in terms of memory consumption and computational complexity and the complexity of the implementation on the sensor nodes.

Instead of collecting vectors of sensor values at a central base station, neighboring sensor nodes exchange sensor readings directly with each other. Based on received data, a sensor node is able to gradually calculate similarity values comparing its own values to the sensor values of its neighbors. Only these similarity values later need to be collected by a central base station which is then able to calculate a clustering of nodes based on this information. Note that the amount of data to be collected is very small compared to collecting whole sets of sensor readings.

Performing large parts of the node clustering procedure distributed in the network requires careful reconsideration of the preprocessing, similarity calculation and data clustering methods. In particular, their calculation must not require storing state on the sensor node that grows with the number of samples collected to account for the scarce amount of memory available on the sensor nodes. For example, k-means clustering – a partitional clustering approach we successfully used in the centralized analysis – cannot be used in our distributed implementation as it needs to recalculate similarities based on the original sensor values in each iteration of the algorithm.

In general, one main challenge for node clustering based on sensor data lies in adapting methods to fit the properties and capabilities of wireless sensor nodes. Herein lies an important contribution of our work.

To compensate for temporary weaknesses of individual criteria (e.g., light sensor recordings at night) and to improve the overall clustering quality, it is important to being able to combine clustering information from different sources. We do this with the help of the average consensus supertree (ACS) [3] method which was originally developed in the field of biology.

III. Preliminary Evaluation

We have implemented our sensor-based clustering of nodes for TinyOS 2.0 running on Tmote Sky sensor nodes and using the built-in light, humidity and temperature sensors. We then evaluated the approach in different indoor scenarios with Tmote Sky sensor nodes placed in five different rooms in each scenario.

The overall results of our experiments were quite positive. A lot of different criteria and combinations of criteria are able to correctly cluster 80% and more of the groups correctly using a relatively small number of samples. Among the different types
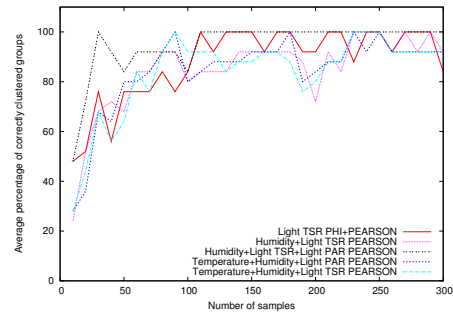

Fig. 2. Percentage of groups clustered correctly for different criteria

of sensors, the light sensors (and here mainly the total solar radiation light sensors) play a particularly important role as information sources. However, the temperature and humidity sensors can also play an important role by balancing weak periods of other sensors and improving the overall clustering result.

Not all clustering criteria worked as well as in the centralized analysis, e.g., clustering based on Euclidean distance between sensor values. Our analyses of this phenomenon suggest that this can be partly attributed to radio irregularities which cause widely varying numbers of data sample pairs to be recorded by different nodes. Other criteria, like the Pearson coefficient, are less susceptible to this effect.

The experiments also showed that combining clustering information from different criteria using ACS is essential to the performance of the algorithm. The quality of these clusterings vary less across different experiments and are generally more stable over time.

To illustrate our findings with an example, Fig. 2 shows the average percentage of groups clustered correctly depending on the number of samples used in a scenario with 15 sensor nodes distributed in five rooms. All shown criteria are based on the Pearson coefficient and combine different sensor values using ACS.

IV. Conclusions

Clustering of nodes is an important technique in wireless sensor networks. Many applications can benefit from doing this clustering based on real-world criteria. We have shown that it is feasible to automatically create clusters that adhere to room boundaries in indoor scenarios by analyzing the measurements of inexpensive and broadly available sensors. We were able to distribute most parts of the algorithm and effectively limit the amount of communication required among the nodes.

References

[1] C. Frank and K. Römer, "Algorithms for generic role assignment in wireless sensor networks," in *Proc. of the 3rd Int. Conf. on Embedded Networked Sensor Systems*, 2005.

[2] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 1999.

[3] F.-J. Lapointe and G. Cucumel, "The average consensus procedure: Combination of weighted trees containing identical or overlapping sets of taxa," *Systematic Biology*, vol. 46, 1997.