Mobile Data Analysis:
From numerical traces to activity recognition

Description:

The constant growth of the number of connected devices, mobile phones and intuitive and natural human-machine interfaces has resulted in the democratization of ubiquitous computing and cyber-physical systems. In such systems, physical and logical sensors provide a huge amount of data that describes users' activity as well as supplementary contextual information. Assuming that context data is a strong indicator of user habits, mining numerical traces, i.e. context data sequences, to extract repetitive patterns should enable to model user activity.

In [2], Rick Moritz has modeled mobile context as a sequence of n-tuples, where n is the number of context data sources, in other words various mobile sensors. Each tuple-element is a discretized sensor reading. An alignment algorithm, similar to [1,5], was also proposed to identify similar subsequences of context, which correspond to routine activity. Unfortunately, the approach resulted with practical and algorithmic issues: 1) the mining provided a series of patterns that occurs twice, without precise information about the exact number of occurrences of patterns; 2) the number of extracted patterns is very large and impossible to manually analyze during an evaluation process and 3) extracted patterns were not exploited in any mobile or off-line application.

To alleviate these issues, we propose to decompose the problem in two steps, similarly to a methodology we have previously applied to find recurrent interaction patterns in a corpus of dialogues, described in [3]. In this work, we have been able to use combinatorial optimization ideas to efficiently solve the two steps and find recurrent interaction patterns that had not previously been detected manually by a human expert or with heuristics. The originality of this work has been recognized by the French operations research society (ROADEF) who gave the Young Researcher Prize to our PhD student in 2015.

For the present work the main steps would be :
1. extraction of interaction patterns from numerical traces,
2. clustering of interaction patterns,
3. exploitation of interaction patterns on a mobile application.

The scientific difficulties are numerous. First of all, the problem modeling would have to take into account a mix of discrete versus continuous sensors on one hand, and logical versus physical sensors on the other hand. Then, the amount of data is far larger than in [3, 4], and we do not expect to obtain good results by using exact optimization methods. Instead, heuristics should be considered, but taking into account the mathematical structure of the problems in order to be efficient. Hence, the PhD should put his effort on approaches related on approximation algorithms and matheuristics (using mathematical programming to design efficient heuristics). Temporal sampling would have to be performed in order to process monthly habits as well as daily ones or short-time series of actions. Finally, interaction patterns can be exploited in various situations: by extracting inter-user similarities, which corresponds to profiling, or by detecting intra-user regularities which corresponds habit detection. In the later case, the mining can be performed off-line, which raises security issues, or on-line, which necessitates very efficient algorithms on mobile devices. Finally, the analysis may benefit from information such as the geolocalization in order to

provide a better extraction of interaction patterns.

Contributions of the PhD student are expected especially to:
- model the problem of regularity extraction from sequences of n-tuples [2,3],
- propose an original interaction pattern extraction method in order to improve the existing ones [1,2,3,5],
- propose an original clustering method of interaction patterns [3,4] based on approximation algorithms and/or matheuristics,
- validate the models on the collected database [2],
- exploit the extracted interaction patterns in a mobile application.

Mobile data was collected during [2] that will be exploited during the thesis.

References:

[1] T. Lecroq, A. Pauchet, E. Chanoni, G. Ayala Solano, *"Pattern discovery in annotated dialogues using dynamic programming"*, International Journal of Intelligent Information and Database Systems, Vol. 6, No. 6, pp. 603-618, 2012.

[2] R. Moritz, *"Routine Activity Extraction from Local Alignments in Mobile Phone Context Data"*, PhD manuscript, 162p., INSA Rouen (supervisors: M. Mainguenaud, A. Pauchet and G. Lefebvre – Orange labs), 2013.

[3] Z. Alès, "Extraction et partitionnement pour la recherche de régularités : application à l'analyse de dialogues. *" (Extraction and partitioning for regularity mining : application to dialog analysis),* PhD manuscript, 180p., INSA Rouen (supervisors: C. Gout, A. Knippel, A. Pauchet and L. Vercouter), 2014.

[4] Z. Alès, A. Knippel, A. Pauchet, *Polyhedral combinatorics of the K-partitioning problem with representative variables*, Discrete Applied Mathematics, pp. 1-14, volume 211/C, 2016.

[5] Z. Alès, A. Pauchet, A. Knippel, *Extraction and Clustering of Two-Dimensional Dialogue Patterns*, International Journal on Artificial Intelligence Tools, Volume 27, N°2, pp. 1-27, 2018.

Supervisors:
Arnaud Knippel (LMI), Bruno Portier (LMI) and Alexandre Pauchet (LITIS).

Laboratories:
LITIS (EA 4108 - FR 3638; http://www.litislab.fr/) is a member of the NormaSTIC (http://www.normastic.fr/) CNRS federation of computer science laboratories of Normandy, while LMI (EA 3226 - FR CNRS 3335; http://lmi.insa-rouen.fr/) is a member of the Normandie Mathematiques (http://normandie.math.cnrs.fr/), the CNRS federation of mathematical laboratories of Normandy.
The two laboratories have been working together for a long time with great results. In particular, the PhD thesis of Z. Ales lead to a 2015 Young Researcher Prize from the national French operational research society ROADEF.

Applicant profile:

The candidate must have a MSc in computer science, operational research or applied mathematics. In particular, expertise and development experiences in the domains of data-mining, machine learning or operational research are required. Some knowledge in artificial intelligence, mobile computing, (geographic) database is also welcome.