Introduction to Parallel Universes and Local Patterns

Michael R. Berthold[1], Katharina Morik[2] and Arno Siebes[3]

[1] University Konstanz, Department of Computer and Information Science
PO Box M712, 78457 Konstanz, Germany
berthold@inf.uni-konstanz.de
[2] University of Dortmund, Computer Science Department, LS VIII
44221 Dortmund, Germany
morik@ls8.informatik.uni-dortmund.de
[3] Utrecht University, Department of Information and Computing Sciences
PO Box 80.089, 3508TB Utrecht, The Netherlands
siebes@cs.uu.nl

1 Introduction

Learning in parallel universes and the mining for local patterns are both relatively new fields of research. Local pattern detection addresses the problem of identifying (small) deviations from an overall distribution of some underlying data in some feature space. Learning in parallel universes on the other hand, deals with the analysis of objects, which are given in different feature spaces, i.e. parallel universes; and the aim is on finding groups of objects, which show "interesting" behavior in some of these universes. So, while local patterns describe interesting properties of a subset of the overall space or set of objects, learning in parallel universes also aims at finding interesting patterns across different feature spaces or object descriptions. Dagstuhl Seminar 07181 on Parallel Universes and Local Patterns held in May 2007 brought together researchers with different backgrounds to discuss latest advances in both fields and to draw connections between the two.

2 Local Patterns

Research on local pattern detection emerged from the fact that most traditional methods in knowledge discovery and databases (KDD) seek to find global models, which describe the overall structure of a dataset and hence explain most of the objects contained in it, but tend to miss local deviations from a background model. The insights learned from such global models are often limited to observations which the domain expert is mostly already aware of and which are therefore not of special interest. In 2002, Hand organized a workshop on pattern detection and discovery and proposed the field of local pattern detection [1]. Since then researchers with different backgrounds (e.g. statistics, machine learning, multi-relational data mining) have come together to establish and unify

Dagstuhl Seminar Proceedings 07181
Parallel Universes and Local Patterns
http://drops.dagstuhl.de/opus/volltexte/2007/1265

the field. Following the 2002 workshop, a second workshop took place in spring 2004 [2] with the goal to find a definition for local patterns. The discussions brought up a number of – often only slightly different – definitions. This seminar continued the inspection of outliers (e.g., Neill Adams, Ira Assent).

3 Parallel Universes

The field of learning in parallel universes originated from the observation that the true objective of *data analysis* is not about mining the data but about mining their underlying objects. These objects are, for instance molecules, images or processes, which (by lack of a better representation) are described based on measurable features (e.g. molecular weight). Such a (set of) features is usually referred to as *data* but obviously there are manifold ways to derive features (or data) while focusing on different aspects of the underlying objects.The notion of learning in parallel universes has first been introduced in [3]. The different feature spaces are regarded as parallel universes and the analysis in parallel is called learning in parallel universes. The aim is to identify configurations in the data, which are shared among different -not necessarily all- universes but also those which are typical to individual universes only. Communities of users which share a certain view of a collection of objects also share a set of features describing the objects, where other communities constitute another view of the same collection [4]. The notion of parallel universes has obvious connections to the research field of Multi-View learning [5]; however, multi-view learning requires all universes (or views) to contain the same information, i.e. there are no patterns specific to individual universes only.

4 How Local Patterns link to Parallel Universes

Throughout the seminar, there were lively discussion as to where to draw a connection between local pattern detection and learning in parallel universes. One obvious link is locality. In terms of local pattern detection it addresses the identification of small deviations from a background model. Similarly, in terms of parallel universes it means the identification of certain patterns that are typical to few (in its extreme one) universes only. Both a single local pattern as well as a pattern which occurs in one/few universes, can give valuable insights to the expert.

Locality in subgroup discovery (Stefan Rüping, Martin Scholz), in term sets (Francesco Bonchi, Jean-Francois Boulicaut, Bruno Cremilleux, Elisa Fromont), in clustering (Michael Berthold, Frank Höppner, Katharina Morik, Bernd Wiswedel), and over time (Bart Goethals, Frank Höppner) was investigated with respect to its link to parallel universes.

A link between both concepts can be drawn by reducing the learning in parallel universes to, without loss of generalizability, learning with different similarity measures. A local pattern induces also a (simple) similarity measure: two tuples are either equal or they are not (Arno Siebes). Similarly, a link was drawn

between parallel universes and multi-objective learning (Ingo Mierswa, Claus Weihs): a universe is constituted by a criterion of success.

5 Organization

The goal of the proposed workshop was threefold. Firstly, we wanted to bring together researchers from the different disciplines to agree on a unifying framework for local pattern mining in parallel universes. So far, only algorithms that find clusters as local patterns have been proposed, for example for the grouping of active molecular compounds or the modeling of user preference clusters in different musical genres. It is not straight forward to extend this scenario to other types of pattern mining algorithms, which requires a careful study of the state of the art and a combination of existing approaches. Secondly, the interaction between different local patterns is an aspect that hinders existing algorithms. If a pattern belongs to two or more local patterns or, inversely, if two local patterns in different universes describe overlapping subsets of the data it becomes more complicated to algorithmically derive the entire set of local patterns that may exist in the data. Thirdly, the workshop aimed to produce a series of white papers describing the state of the art in local pattern mining in application areas where related problems have appeared in the past.

In order to achieve this, we invited researchers from different communities: local pattern mining, statistical data analysis, machine learning, and data mining. In addition we also invited participants from the Visual Data Mining community, since local pattern detection -especially in several descriptor spaces in parallel- is a method that inherently requires user feedback to be successful (Rudolf Kruse, Matthias Steinbrecher). For this, it is crucial to be able to present the user with a variety of -preferably interactive- views on the data (Arno Knobbe), each showing summaries of the discovered patterns in each universe together.

References

1. Hand, D.J., Adams, N.M., Bolton, R.J., eds. In Hand, D.J., Adams, N.M., Bolton, R.J., eds.: Pattern Detection and Discovery (ESF Exploratory Workshop, London, UK, September 16-19, 2002). Volume 2447 of Lecture Notes in Computer Science., Springer (2002)
2. Morik, K., Boulicaut, J.F., Siebes, A.: Local Pattern Detection: International Seminar Dagstuhl Castle, Germany, April 12-16, 2004, Revised Selected Papers (Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2005)
3. Patterson, D.E., Berthold, M.R.: Clustering in parallel universes. In: IEEE Conference on Systems, Man and Cybernetics, IEEE Press (2001)
4. Wurst, M., Morik, K., Mierswa, I.: Localized alternative cluster ensembles for collaborative structuring. In Fürnkranz, J., Scheffer, T., Spiliopoulou, M., eds.: Proceedings of the European Conference on Machine Learning, Berlin, Springer (2006) 485–496

5. Rüping, S., Scheffer, T., eds.: Proceedings of the ICML 2005 Workshop on Learning With Multiple Views. http://www-ai.cs.uni-dortmund.de/MULTIVIEW2005/MultipleViews.pdf (2005)