

MIME: A Framework for Interactive Visual Pattern Mining

Bart Goethals, Sandy Moens, and Jilles Vreeken

University of Antwerp, Belgium

Abstract. We present a framework for interactive visual pattern mining. Our system enables the user to browse through the data and patterns easily and intuitively, using a toolbox consisting of interestingness measures, mining algorithms and post-processing algorithms to assist in identifying interesting patterns. By mining interactively, we enable the user to combine their subjective interestingness measure and background knowledge with a wide variety of objective measures to easily and quickly mine the most important and interesting patterns. Basically, we enable the user to become an essential part of the mining algorithm. Our demo currently applies to mining interesting itemsets and association rules, and its extension to episodes and decision trees is ongoing research.

Keywords: MIME, Pattern Exploration, Interactive Visual Mining

1 Introduction

Data mining is an inherently iterative process; the results of one analysis often lead to new questions, requiring more analysis. In an ideal world, this process is streamlined. That is, data mining is not only iterative, but also interactive: the user can give such feedback immediately, and easily browse the results. In traditional pattern mining, however, algorithms typically produce large amounts of patterns, many of which are not interesting to the user [8], and the results are typically only given in a flat text file, making it hard to analyze the results. By instead providing an iterative and interactive process, the user would be able to explore and refine the discovered patterns on the fly.

In this demo, we present a framework in which we allow the user to interactively mine a database for interesting itemsets and association rules. Our system visualizes all patterns discovered so far, yet, importantly lets the user interactively explore and dynamically modify these in an intuitive manner [2, 9]. By the visualization, users can browse through the mined data more easily and so quickly discover important information. Essentially, in our framework visualization strengthens the interactive mining process, and vice versa.

Basically, our framework, MIME (Making Interactive Mining Easy), draws from user knowledge and interest to improve the collection of patterns discovered by the mining algorithms, by letting the user take control during the mining

process, and allowing to adapt the results and so create useful collections of patterns. To assist the user, MIME offers an extensive toolbox of interestingness measures, mining techniques and visualizations. Using these, the user can easily identify and remove uninteresting or redundant patterns, extend or reduce existing patterns, or apply various post-processing techniques. As such, in our framework the user becomes an essential part of the mining algorithm.

2 Description of the System

We consider transactional (supermarket) databases where a transaction contains a number of items. Here, a pattern is an itemset or an association rule [7].

Most pattern mining techniques produce an amount of output that due to size is difficult to post-process. One could try to reduce the number of results by making the quality thresholds more strict. Unfortunately this does not guarantee the usefulness of the produced patterns. To this end, more is required.

MIME combines the knowledge of the expert and the computational strength of a computer to increase the probability of finding interesting patterns. This is achieved in a visual framework where the expert can create his own pattern collection. In order to evaluate the created/mined patterns, MIME implements many interestingness measures. This way the user can select a number of measures that are important for his purpose/domain. These measures improve the knowledge of the user. Our tool also contains a number of standard mining and post-processing algorithms such as Eclat and Apriori. The mining algorithms can be used to construct a starting set of patterns a user can play with and can further construct his collection of useful patterns from.

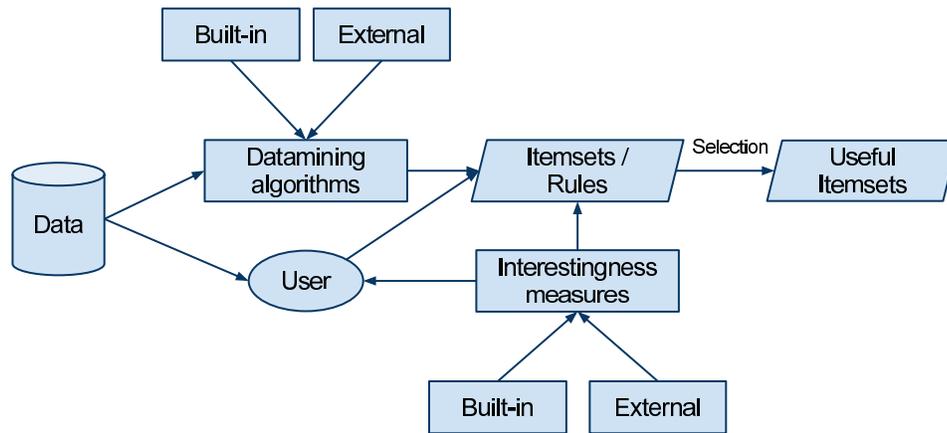


Fig. 1. MIME Workflow

The overall workflow of MIME is shown in Figure 1. Data is read from a database file and can be used by a user or by mining algorithms to create rules

and itemsets. From the created patterns a user can make a decision which patterns are useful, based on the provided interestingness measures. The user may also apply post-processing algorithms to the created collection of patterns.

An important aspect of MIME is the way in which the user interacts with the current state of the system. All interactions made by the user have a direct impact on its current state (i.e. the items, itemsets and association rules). Thereto MIME contains 4 different panes a user can interact with: the *source dock*, the *work dock* (or *workspaces*), a *toolbox* and the *global overview*. The source dock visualizes items in the dataset. All items are ranked in decreasing order based on an active interestingness measure. The work dock shows mined patterns. Patterns can be altered, deleted, expanded, etc. The toolbox contains measures, mining algorithms and post-processing algorithms which can be applied. The global overview provides more general information about the workspaces.

All information provided by MIME is computed on-the-fly. When selecting new measures from the toolbox, it computes the corresponding ranks of the discovered patterns and shows them as soon as they are available. Extending or reducing patterns also reflects in an immediate recalculation of the applied measures. The nature of any-time algorithms applies to our framework in the sense that partial information is shown when available. Therefore, we have made extensive use of caching and threading in our system. Using caching, MIME minimizes the number of computations that are made by reusing previously computed information. Threading is used to compute information in the background.

An important feature is the *Best Pattern Extension*. This feature allows the user to immediately see the individual impact of remaining items on an existing pattern, i.e. a pattern is repeatedly extended by one of the remaining items in the dataset and the ranks (based on the active measure) are computed and shown in the source dock, again using rank decreasing order. Other functionalities for generating patterns based on existing patterns, are the generation of all subsets, subsets of given size, closed supersets, etc. Mining algorithms can also use discovered patterns to filter the mining output. All these features (including mining and post-processing algorithms) provide the user with a toolbox for easy and fast generation of possibly useful patterns.

From the workflow notice the *External* nodes indicating the use of external/plugin functionality in the tool. In order to provide a widely applicable and easily extensible tool, we have equipped our tool with a plugin system, such that existing mining and post-processing algorithms can be used. The produced patterns are automatically loaded into the MIME framework.

A demonstration video introducing the basic functionalities of MIME can be found on the webpage of our research group¹.

3 Related Work

A lot of work has been done comparing and evaluating different objective interestingness measures [6, 10]. In our tool several objective interestingness measures

¹ <http://adrem.ua.ac.be/mime>

have been incorporated, but it is the combination of user-knowledge and objective measures enabling subjective interestingness criteria to be applied.

Also in the context of Inductive Databases several interactive constraint-based mining frameworks have been studied [3]. Our framework could be built on top of such an inductive database implementation.

Most similar to the system presented here, Ankerst et al. proposed a framework for mining decision trees [1]. The user and computer work together in this process such that in each step either the user can make the decision for a new split or the computer can make this decision. The computer also provides extra computational power by showing the best split, look-ahead information, purity measures etc. Our approach is similar, but applies to frequent itemsets and association rules, instead of decision trees.

Some methods for visualizing patterns exist [4, 5]. These visualizers present the output of mining algorithms in a compact and graphical format, and allow to further filter the output using queries. They do not provide means to mine the database interactively using subjective criteria and also do not allow to further explore existing patterns.

References

1. Mihael Ankerst, Martin Ester, and Hans-Peter Kriegel. Towards an effective co-operation of the user and the computer for classification. In *Proc. ACM SIGKDD*, pages 179–188, 2000.
2. Enrico Bertini and Denis Lalanne. Investigating and reflecting on the integration of automatic data analysis and visualization in knowledge discovery. *SIGKDD Explor. Newsl.*, 11:9–18, May 2010.
3. Hendrik Blockeel, Toon Calders, Elisa Fromont, Bart Goethals, Adriana Prado, and Céline Robardet. An inductive database prototype based on virtual mining views. In *Proc. ACM SIGKDD*, pages 1061–1064, 2008.
4. Christopher L. Carmichael and Carson Kai-Sang Leung. FpVAT: a visual analytic tool for supporting frequent pattern mining. *SIGKDD Explor. Newsl.*, 11:39–48, May 2010.
5. Amitava Datta and Kesaraporn Techapichetvanich. VisAR : a new technique for visualizing mined association rules. In Xue Li, Shuliang Wang, and Zhao Dong, editors, *Adv. Data Min. Appl.*, pages 728–728. Springer Berlin / Heidelberg, 2005.
6. Liqiang Geng and Howard J. Hamilton. Interestingness measures for data mining: A survey. *ACM Comput. Surv.*, 38, September 2006.
7. Bart Goethals. Frequent set mining. In *Data Mining and Knowledge Discovery Handbook*, pages 321–338. 2010.
8. Bart Goethals, Nikolaj Tatti, and Jilles Vreeken. Useful patterns (UP’10) ACM SIGKDD workshop report. *SIGKDD Explor. Newsl.*, 12:56–58, March 2011.
9. Daniel A. Keim. Information visualization and visual data mining. *IEEE Trans. Vis. Comp. Graph.*, 8:1–8, January 2002.
10. Vipin Kumar, Jaideep Srivastava, and Pang-Ning Tan. Selecting the right interestingness measure for association patterns. In *Proc. ACM SIGKDD*, pages 32–41, 2002.