



Visualisation for Data Mining telecommunications network data

R. Sterritt, E.P. Curran, K. Adamson, C.M. Shapcott
Faculty of Informatics, University of Ulster, Northern Ireland.

Abstract

It may be proposed that a flaw in data mining is that it is not user-centred. It would be helpful to visualise the data at all stages to enable the user to gain trust in the process and hence have more confidence in the mined patterns. The transformation from data to knowledge requires interpretation and evaluation, which also stands to benefit from multi-stage visualisation of the process. This paper discusses how a knowledge discovery architecture developed for Telecommunications Management Network (TMN) data was extended with visualisation tools throughout all stages of the process to facilitate the aim of making the data mining application more user-centric.

1 Introduction

It may be proposed that a flaw in data mining is that it is not user-centred. It would be helpful to visualise the data at all stages to enable the user to gain trust in the process and hence have more confidence in the mined patterns. The transformation from data to knowledge requires interpretation and evaluation, which also stands to benefit from multi-stage visualisation of the process.

This paper discusses how a knowledge discovery architecture developed for Telecommunications Management Network data (Sterritt [1]) was extended with visualisation tools throughout the process to increase the human interaction in the

data mining task. Firstly the interpretation of data mining versus knowledge discovery is presented

1.1 Knowledge Discovery in Databases

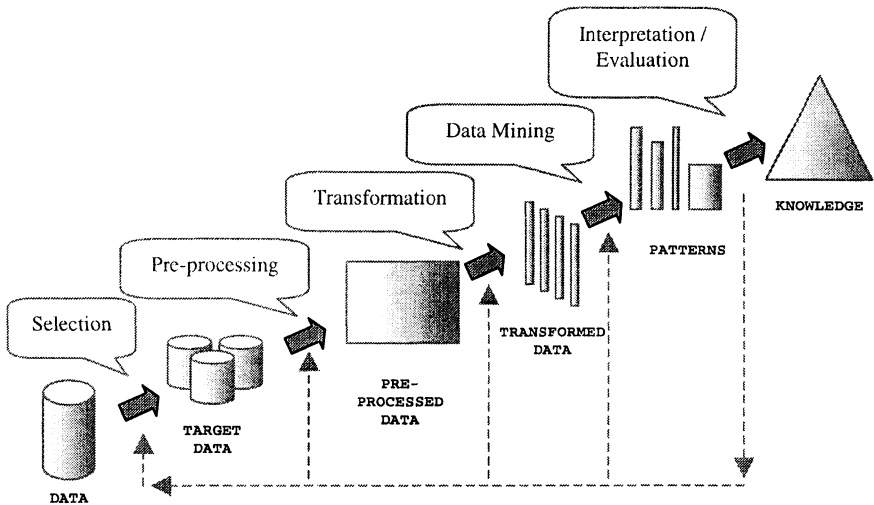


Figure 1 The Knowledge Discovery Process

Knowledge Discovery in Databases (KDD) is considered to be “the non-trivial extraction of implicit, previously unknown, and potentially useful information from data” (Frawley [2]). This implies a focus only on the discovered information, yet the current opinion is that KDD means more than this. KDD refers to the overall process of discovering useful knowledge from data, while data mining refers to the application of algorithms for extraction purposes (Fayyad [3]). Brachman & Anand [3][4] present a process that includes human intervention (Figure 1). Although autonomous KDD may be desirable in the long run this is not the current state of affairs [4]. It has therefore been highlighted that KDD researchers need to place more emphasis on the overall KDD process and on tools to support its various stages (Uthurusamy [5]).

2. The Visualisation Tools

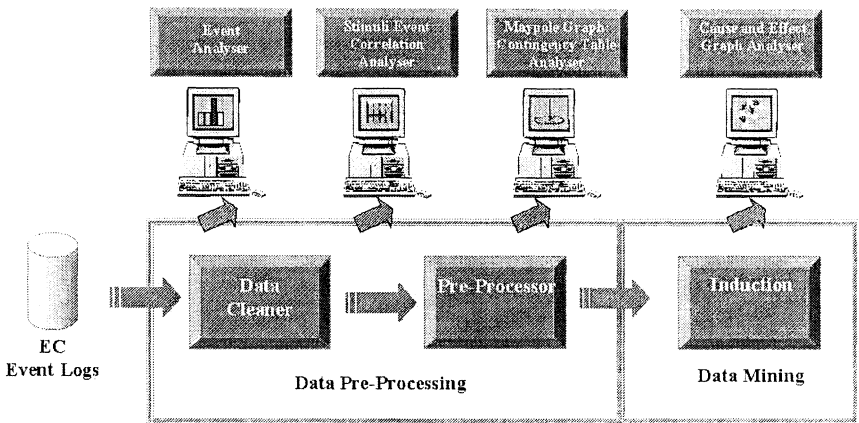


Figure 2 Visualisation of the Knowledge Discovery Process

It is prevalent in the literature that upon engaging in real-world discovery tasks it has been found that they can be extremely complex [4]. Adding visual tools to the process can reduce this complexity by facilitating understanding of the data and patterns discovered.

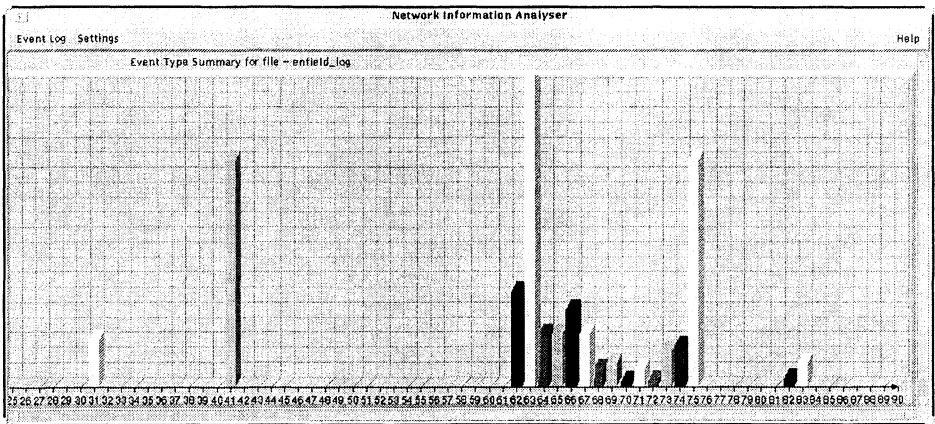
Visualisation applied to KDD can offer 'human-assisted computer discovery' and 'computer-assisted human discovery'. Such a visual environment, by reducing the time to understand complex data, would enable practical solutions to many real world problems to be developed far more rapidly than either human or computer operating independently [5]. In doing so the remarkable perceptual abilities that humans possess can be utilised, such as the capacity to recognise images quickly, and detect the subtlest changes in size, colour, shape, movement or texture [6].

The tools that were developed for this exemplar to allow the visualisation of the knowledge discovery process are as follows (Figure 2):

- ☐ Event Analyser
- ☐ Stimuli-Event Correlation Analyser
- ☐ Contingency Table Analyser
- ☐ Cause and Effect Graph Analyser

These tools were mostly developed in Java to facilitate both Unix and PC environments.

2.1 Event Analyser



Screenshot 1 The Event Analyser

Visualisation techniques are a very useful method of discovering patterns in data sets, and may be used at the beginning of the data mining process to get a rough feeling of the quality of the data set and where patterns are to be found (Adriaans [7]).

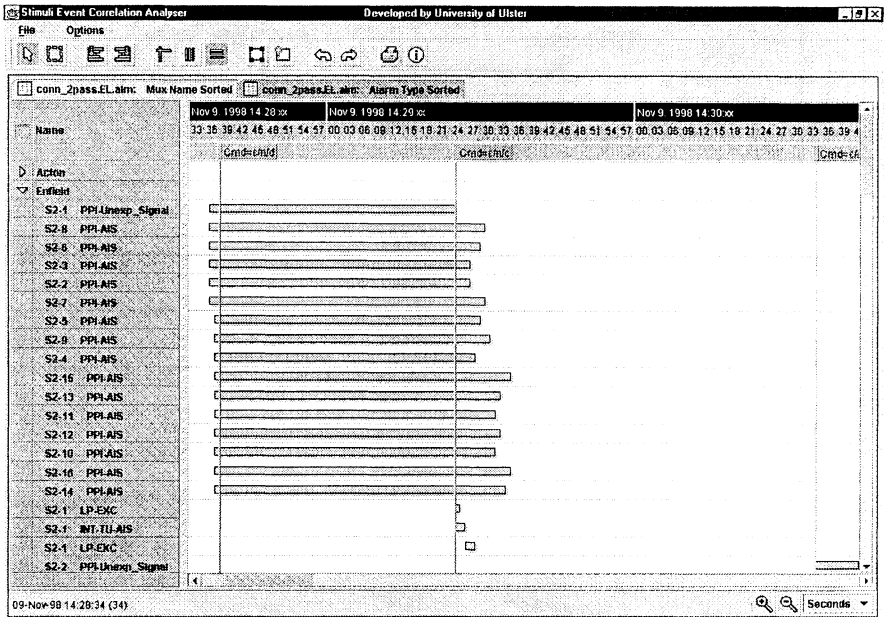
A bar chart or histogram is one of the simplest and easily understood charts available. It is now so common that with the use of wizards it can be drawn from data in databases with a small number of user mouse clicks.

This Java bespoke version (Screenshot 1 The Event Analyser) is a part of the data cleaning process. It displays the frequency of alarms in the data that is being cleaned. As such it can be used to judge the quality of the data; for instance the lack of alarms in the range 43-61 (centre of the screenshot) may indicate that the data set is not adequate.

The fault data is produced in the testing environment as well as for management of operational networks. Since test cases are often reused, the tool can be used for user pattern recognition to give a rough indication of a pass or fail compared with previous test runs i.e. computer assisted human discovery.



2.2 Stimuli-Event Correlation Analyser



Screenshot 2 Stimuli Event Correlation Analyser

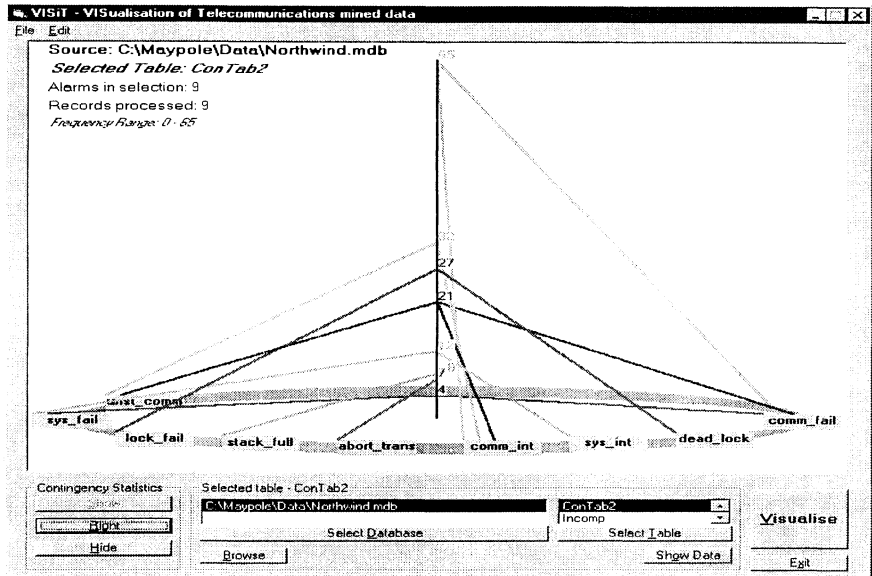
The next tool allows the rapid and concise visualisation and navigation of extensive telecommunications data, specifically the event log from the Element Controller. Traditionally this source of historic data was not utilised because of the huge mountain of data generated.

The tool, Screenshot 2 Stimuli Event Correlation Analyser, (Sterritt [8]) displays the user actions (vertical lines) against the alarm data (horizontal bars) in an easy to read colour coded 'Gantt' chart. Alarms represent the symptoms of a fault, which may occur naturally in an operating network or as a result of a user action.

This application not only allows quick data visualisation of the domain that is being mined by the knowledge discoverer but also the test engineer where the user actions are equivalent to the test commands form a test script.



2.3 Contingency Table Analyser



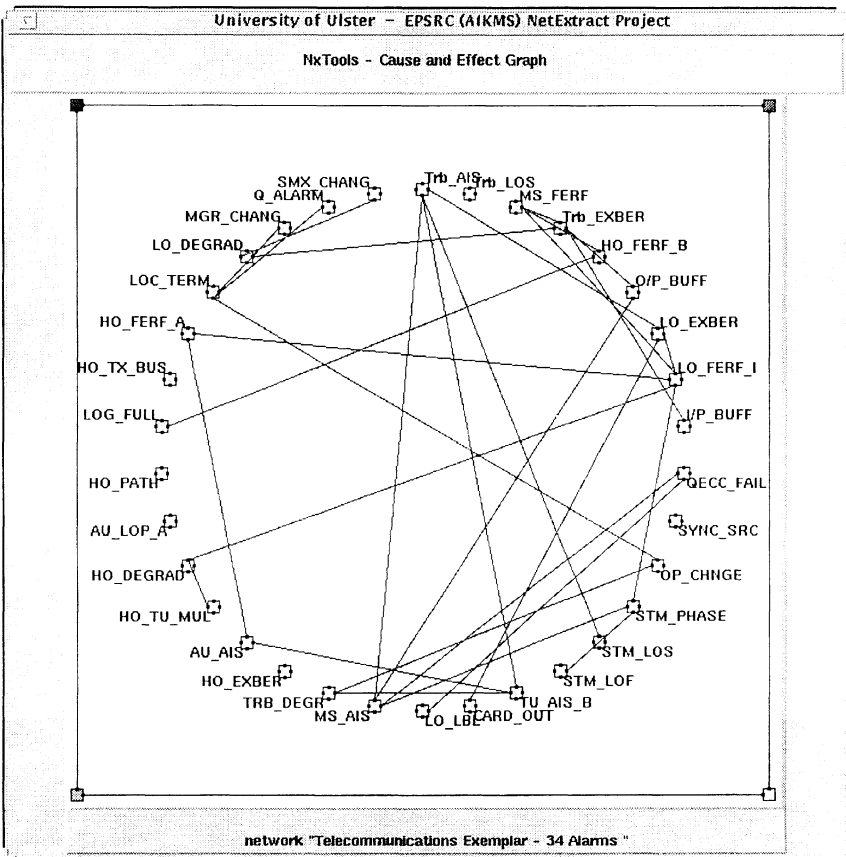
Screenshot 3 The Contingency Table Analyser

The next stage in making the knowledge discovery architecture more user-centric involved visualising the contingency tables. A contingency table, in simple terms, is a list of various variables and their states in combination that have occurred in the history observed, and the frequency of occurrence. It is the final stage of the pre-processing before the actual mining.

Several possibilities were designed for visualising the contingency table; variable-pair histogram, parallel co-ordinate plotting, and the web model. The implemented design allows users to select data sources and visualise the data in a custom graphic referred to as the "Maypole Graph", Screenshot 3 The Contingency Table Analyser, (McBride [9]). All relationships can be visualised by assigning a different colour to each row in the contingency table and showing the links passing from all alarms to a central point (like ribbons hanging from a maypole). By doing this, the user can see all triple and quadruple links as well as the binary links.



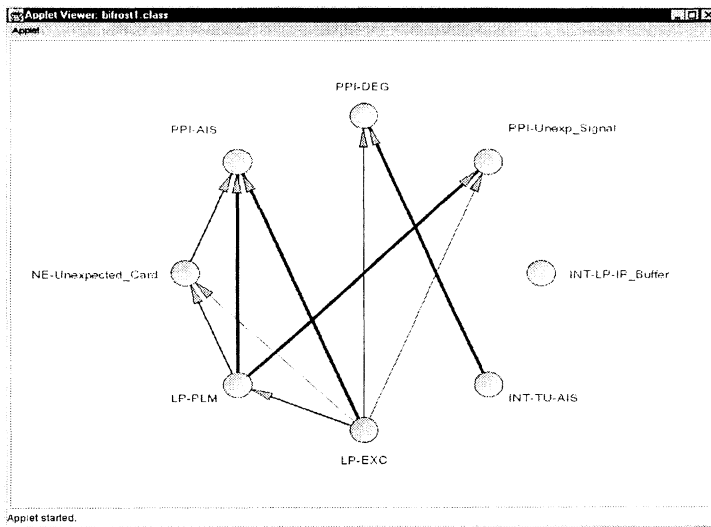
2.4 Cause and Effect Graph Analyser



Screenshot 4 Cause and Effect Graph Analyser

The last visualisation tool was a part of the original architecture since the mined results needed to be visualised. Is this application of KDD the mining algorithms developed (Sterritt [10], Shapcott [11]) produce a probabilistic network.

The tool has been rewritten and now also included the strengths of connection between the variables (in this case the alarm events).



Screenshot 5 Probabilistic Network Analyser

3. Conclusion

3.1 Evaluation

This work has produced a useful set of visualisation tools to assist in making a Knowledge Discovery architecture more user-centric at all stages of the process. To achieve discovered knowledge from data requires “interruption and evaluation”, the human element, these tools facilitate this aim.

The tools use simple graphical display techniques, which provide a wealth of information and assist in the understanding of the process, however, the use of three-dimensional techniques would allow better navigation and provide a better context of the vast amount of data. Figure 3 highlights the limitations of providing a proper context with two-dimensional techniques.

Some of the tools have proved to be useful as stand-alone applications, in particular the Gantt tool which allows rapid visualisation of the Element Controller’s event log.

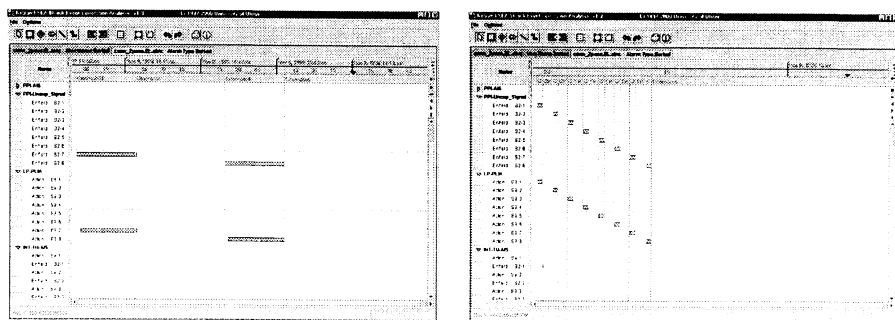


Figure 3 Limited Context in the Gantt tool

3.2 Future Work

With the aim of increasing the user-centric nature of the architecture it is planned that the tools be developed beyond visual aids. They will become more integrated with the actual process and allow the user to interactively work through the process. For example the Gantt tool could allow the user to pick the area of data (events and/or timeframe) to be mined.

Acknowledgements

We are greatly indebted to our industrial collaborators Northern Ireland Telecommunications Engineering Centre (NITEC), Nortel Networks, who have supported our research for many years now, and for their faith in direct funding the latest project (Jigsaw programme 1999-2002). We would also like to thank EPSRC (AIKMS programme 1995-97), IRTU (Start programme 1996-99) for funding the research of which this was a sub-project.

References

- [1] Sterritt, R., Adamson, K., Shapcott, C.M., Bell, D.A., An Architecture for Knowledge Discovery in Complex Telecommunication Systems, eds. Adey R.A., Rzevski G., Nolan P., *Applications of Artificial Intelligence in Engineering XIII*, Computational Mechanics Publications: Southampton, CD-ROM pp627-640, 1998
- [2] Frawley, W., Piatetsky-Shapiro, G., Matheus, C., Knowledge Discovery in Database: An Overview. *AI Magazine* 14(3): pp57-70, 1992
- [3] Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. From Data Mining to Knowledge Discovery: An Overview, *Advances in Knowledge Discovery & Data Mining*, AAAI Press & The MIT Press: California, pp1-34 1996



- [4] Brachman, R.J., Anand, T., The Process of Knowledge Discovery in Databases: A Human-Centered Approach., *Advances in Knowledge Discovery & Data Mining*, AAAI Press & The MIT Press: California, pp37-57, 1996
- [5] Uthurusamy, R. "From Data Mining to Knowledge Discovery: Current Challenges and Future Directions", *Advances in Knowledge Discovery & Data Mining*, AAAI Press & The MIT Press: California, pp 561-569, 1996
- [6] <http://www.avis.com/solution/>, Advanced Visual Systems Incorporated.
- [7] Adriaans, P., Zantinge, D., *Data Mining*, Addison-Wesley: Harlow, England, 1996.
- [8] Sterritt, R., Adamson, K., Curran, E.P., Shapcott, C.M., Visualisation And Context Of Telecommunications Data, *Proceedings of the 17th International IASTED Conference on Applied Informatics*, 1999
- [9] McBride, S., Sterritt, R., Curran, E.P., Adamson, K., Shapcott, C.M., "MAYPOLE: Visualising Contingency Tables", *Accepted for the 18th International IASTED Conference on Applied Informatics*, 2000
- [10] Sterritt, R., Adamson, K., Shapcott, C.M., Bell, D.A., McErlean, F., Using A.I. For The Analysis Of Complex Systems, *Proceedings of the IASTED International Conference on Artificial Intelligence and Soft Computing* pp113-116, 1997
- [11] Shapcott, C.M., Sterritt, R., Adamson, K., Curran, E.P., NETEXTRACT - Extracting Belief Networks in Telecommunications Data, *Proceedings of the ERUDIT Workshop on Application of Computational Intelligence Techniques in Telecommunication*, 1999