International Symposium on Visual Analytics Science and Technology (2010)
J. Kohlhammer and d. Keim(Editors)

Does Jason Bourne need Visual Analytics to catch the Jackal?

A. Bertone, T. Lammarsch, T. Turic, W. Aigner, S. Miksch

Department of Information and Knowledge Engineering, Danube University Krems

ABSTRACT

Visual Analytics is a relatively new field which tries to combine and intertwine visual and analytical methods in an interactive manner. Because of the complex structure of time, the application of visual analytics methods to time-oriented data is a very promising approach for insight generation. To show how this can be applied, on top of real world data we created a fictitious scenario where even one of Ludlum's heroes, Jason Bourne, could take advantage of the collaboration between visual and analytical methods.

Categories and Subject Descriptors: H.2.8 Database Applications: Data mining, I.3.6 Methodology and Techniques: Interaction Techniques, I.5.2 Design Methodology: Pattern analysis

1. Introduction

"Bourne hung up the phone and returned to the couch and the printouts, separating three that had caught his attention, not that any of them contained anything that evoked the Jackal. Instead, it was seemingly offhand data that might conceivably link the three to each other when no apparent connection existed between them. According to their passports, these three Americans had flown in to Philadelphia's International Airport within six days of one another eight months ago. Two women and a man, the women from Marrakesh and Lisbon, the man from West Berlin. The first woman was an interior decorator on a collecting trip to the old Moroccan city, the second an executive for the Chase Bank, Foreign Department; the man was an aerospace engineer on loan to the Air Force from McDonnell-Douglas. Why would three such obviously different people, with such dissimilar professions, converge on the same city within a week of one another? Coincidence? Entirely possible, but considering the number of international airports in the country, including the most frequented-New York, Chicago, Los Angeles, Miami - the coincidence of Philadelphia seemed unlikely. Stranger still, and even more unlikely, was the fact that these same three people were staying at the same hotel at the same time in Washington eight months later[1]".

This short excerpt from Ludlum's Bourne Ultimatum describes a critical situation that analysts and decision makers usually have to face. A huge amount of time related information apparently not connected with one another, many sources in different format, collected together in a not usable manner and different kind of data.

In many application fields the analysis of such data can take advantage of a relatively new research field called Visual Analytics. It is defined as "*the science of analytical reasoning facilitated by interactive visual interfaces*" [TC05]. As a matter of fact, combining and intertwining analytical and visual abilities in an interactive manner would improve the analytic power of the methods available to date.

Our focus is in particular on time-oriented data: problems involving time-oriented data need special attention since these data are different from other kinds of data. The reason is that time has an inherent structure, such as calendar aspect being composed of smaller granularities, like seasons and years. Yet, natural and social aspects are deeply influenced by these granularities; therefore, the adoption and exploitation of the structures of time in data analysis methods can massively improve the amount of information gained.

To this aim, we present a ease to use Visual Analytics application in this paper, which can help Ludlum's hero as well as analysts and decision makers to discover useful insight in data, by explicitly using the richer structure of time. The prototype we present is still in work-in progress, but has been designed with the aim to be open to both new visual and analytical methods, which can improve its capabilities.

2. Related work

As stated by [RFP09], achieving the level of fully integrated visualization and automated analysis is a very

[1] R. Ludlum - The Bourne Ultimatum (1990) – ISBN 0394584082

prominent requirement for Visual Analytics. Starting from a bibliographic research to categorize techniques, trends, gaps and potential future directions, [BL09] provide an overview of the integration effort between the aspects of automated analysis and visualization. They show that there are different levels of integration that often lean towards one of the two aspects. This results in either a process for automated analysis or a process for visualization. They also propose potential extensions and research questions to further advance and integrate these fields. To this regard, some attempts to integrate these abilities already exist in literature. For instance, some of them focus on the analysis of time series by using tree visualizations and interactions (e.g., VizTree [LKL*04]); or propose a combination of Visual Data Mining and time series (e.g., Parallel Bar Chart [CCT03]), or combine Data Mining concepts and visualizations (e.g., Statigrafix[2], HCE [SS05]). However, some limitations are still to be faced with, such as the lack in the visualization part, or the need of a strong expertise in the application field, as well as the mining task left to the user. Moreover, some Visual Data Mining and Visual Analytics tools are already available. For example, KNIME[3], Weka[4] and Rapid Miner[5] are representative cases of the former group, whereas Tableau[6] and Spotfire[7] are well known examples of the latter one. However, as it was also outlined in the recent VAKD09[8] workshop, the relevance of a combined use of these two aspects is emphasized. In particular, the need for visually controllable automatic methods, that is, algorithms that are fast enough for visual interaction and whose model structure can be represented visually and controlled using visual interaction.

3. Problem definition

The excerpt from Ludlum’s Bourne Ultimatum provided a quite interesting though fictitious example of a situation where a Visual Analytics approach may be advantageous. As [Ber09] outlined, most of the analytical methods (i.e., sequence and interval mining methods) which try to find interesting patterns from time-oriented data, usually give as result a sequence of events, lacking any knowledge either about the intervals between them or about after how much time a particular pattern will reoccur. Moreover, they usually do not involve the user in the analysis, but rather provide a sort of black box to be applied, neglecting the possibility to add any user knowledge. A less fictitious example which clarifies the problem is shown in Figure 1. A customer decides to buy a Play Station 3 (PS3) gaming console (event A). The next day, s/he buys a racing game (event B). After five days, s/he buys a steering wheel controller (event C). Another racing game is bought after 3 months (event D). If we take into account only the sequence or the order of these events, ABCD, we cannot know after how much time the next item will be purchased. Moreover, we cannot even know after how much time a similar sequence will occur. On the contrary, if also the time intervals are considered, we can not only profile the users according to their interests, habits and requirements, but we can also improve the selling strategies according to the timing of their shopping habits. As a matter of fact, the webshop can vary its offers and catalogues according to the users. For instance, it is possible to send e-mails or letters describing discounts on games for PS3 two months after the first purchase, or make special offers dedicated to those who bought a PS3 in the previous two/three months.

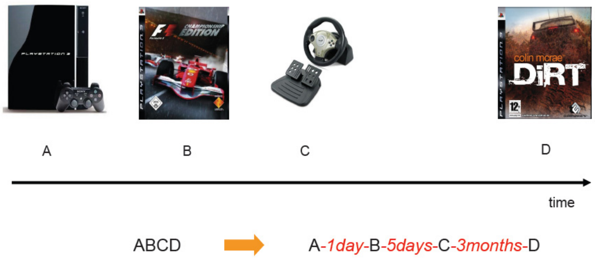


Figure 1: *Event Sequence. Left: Traditional sequence pattern (order only); Right: Multi-time interval pattern (also the intervals between events are provided).*

Beside that, most of the traditional methods rarely provide a visualization of the whole analysis process, and even in case they do, it is often a sort of presentation of the results, rather than an interactive way to further improve the analysis itself. To overcome the first part of these limitations, [Ber09] proposed the use of so called multi-time interval patterns as well as a novel approach to preserve temporal information in between. Moreover, he sketched the possibility to define a visual analytics framework, which is able to bring together the analytical and the visual abilities in an interactive manner. An example in the next section will explain this approach as well as the first stages of the prototypical implementation.

3.1 Application example

Let’s imagine that through his well established network at the CIA in Langley, Alex Conklin is able to find some interesting information about the travels of the Jackal’s *old men*[9]. Let’s suppose the data contain information about number of passengers (economy and business class), depar-

[2] http://statigrafix.com/ (accessed on 18/02/2010)

[3] http://www.knime.org/ (accessed on 18/02/2010)

[4] http://www.cs.waikato.ac.nz/ml/weka/ (accessed on 18/02/2010)

[5] http://rapid-i.com/ (accessed on 18/02/2010)

[6] http://www.tableausoftware.com/ (accessed on 18/02/2010)

[7] http://spotfire.tibco.com/Solutions/Manufacturing-Analytics/ (accessed on 18/02/2010)

[8] http://www.hiit.fi/vakd09/ (accessed on 18/02/2010)

[9] In Ludlum’s books, Carlos “the Jackal” uses old men as couriers throughout Europe for his affairs and as killers. As Jason Bourne outlines “*Who suspects decrepit old men, whether they're beggars or whether they're just holding on to the last remnants of mobility?*

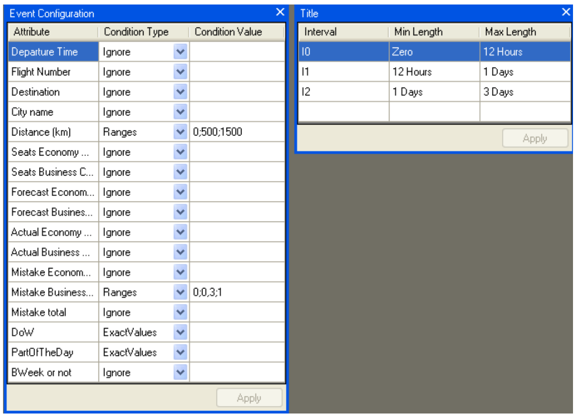
ture time, flight number, flight destination, and number of seats on the flight (both economy and business). Moreover, they also provide information about the number of expected suspected old men for any flight and day (economy and business), according to a not given CIA forecast algorithm, whose forecast are provided about 2 months in advance and then further released six, four, and two weeks before the scheduled departure. Alex Conklin then sends these data to his friend Jason Bourne, who has to derive useful insights from them in order to foresee the future strategies of the Jackal and possibly catch him. To this aim Ludlum's hero may adopt our simple Visual Analytics framework.

First of all, he performs a pre-processing phase to introduce some details which may result relevant for the analysis: he adds the distance between departure and destination, and a measure of how good the forecast is. This measure expresses the difference between the expected suspected old men, and the actual number of those really identified as old men (for a certain flight and date and both flight classes). Moreover, he divides each day in Morning (from 7 till 12), Afternoon (from 12 till 19), Evening (from 19 till 24) and Night (from 0 till 7) as well as the possibility to distinguish between businessweek (Monday to Friday) and weekend (Saturday and Sunday).

After this phase, he can configure the input parameters: as shown in the "*Event Configuration*" (Figure 2, left): he classifies (discretization) the destinations into short distance (0-500 km) and middle-distance destinations (500-1500 km); then takes the part of day as an exact value as well as the day of the week, and the absolute error is divided into small error (from 0 to 0.3) and large error (from 0.3 to 1). Moreover, using the given granularities (i.e., milliseconds, seconds, minutes, hours, days, weeks, months, quarters, and years) he provides three different intervals: from 0 to 12 hours, from 12 hours to 1 day and from 1 day to 3 days respectively (Figure 2, right).

Each step of the analysis process as well as most of the interaction are managed using "2C", a concentric circles visualization ("*to see*"): the chosen intervals are represented as concentric circles, starting from the innermost one (I_0); each event is represented as a bubble in the centre of the concentric circles, whose size is proportional to the number of occurrences of the event itself. Thus, the higher number of times an event occurs, the bigger is the bubble representing it. Multi-time interval patterns of length 1 (i.e., patterns composed of two events) are represented using a segment connecting the events in the centre of the circles and events located on the interval/circle after that they occur. The thickness of the connecting segment corresponds to the number of occurrences of the pattern itself. Hence, the thicker the segment is, the higher number of times the pattern occurs. The same representation holds for multi-time interval patterns of length greater than 1.

Jason Bourne can now have a complete view of the first step of the analysis (Figure 3 Left). For instance, the selection the event e_{46} outlines that it occurs 25 times, on Sunday and on the 4th, 11th, 18th, 25th, 28th day of the month. Moreover, using the right click of the mouse or the proper menu, a table will show further information about the selected event (Figure 3 Right). In this case, Jason Bourne may note that these event are related to middle-distance destinations, has an error greater than 0.3 and occur on Sunday afternoon.



Attribute	Condition Type	Condition Value
Departure Time	Ignore	
Flight Number	Ignore	
Destination	Ignore	
City name	Ignore	
Distance (km)	Ranges	0;500;1500
Seats Economy ...	Ignore	
Seats Business C...	Ignore	
Forecast Econom...	Ignore	
Forecast Busines...	Ignore	
Actual Economy ...	Ignore	
Actual Business ...	Ignore	
Mistake Econom...	Ignore	
Mistake Business...	Ranges	0;0,3;1
Mistake total	Ignore	
DoW	ExactValues	
PartOfTheDay	ExactValues	
BWeek or not	Ignore	

Interval	Min Length	Max Length
I0	Zero	12 Hours
I1	12 Hours	1 Days
I2	1 Days	3 Days

Figure 2: *Event Configuration window (left) allows to choose the attributes of interest in order to define the events (e.g., Day of the week, part of the day) and possibly to discretize one or more of them (e.g., Distance, Mistake Business); the Interval window (right) lets the user define the intervals within which to look for patterns.*

Then he proceeds till Step 2 and compare the insights from the economy and the business case. The first main outcome is that the most occurring multi-time interval pattern occurs between Friday and Sunday (then in what in the social time is usually know as "long weekend"). However, firstly they differ in the temporal occurrence of this pattern, that is, Friday morning-Friday afternoon-Sunday afternoon for the economy class and Friday afternoon-Saturday morning-Sunday afternoon for the business class. Secondly and most interestingly, they differ in the absolute error: while for the economy class it is a small one, for the business class it is greater than 0.3. Figure 4 and Figure 5 show this situation in the case of business class.

On top of such results, Jason Bourne conducts a further analysis. Since it may be relevant to know whether the error between the forecast and the actual number of identified old men is negative or positive, that is, whether there were more or less identified than expected, he decides to add such information to the analysis. Moreover, he divides the day into *Business Morning* (from 6 to 9), *Business Evening* (from 16 to 24), and so called *Tourist time* (from 9 to 16), to distinguish between business trips and leisure trips. In this case, the most occurring patterns related to the old men in economy class occur not only during the weekend (or long weekend), but also during the business week. However, the error is always limited, assuming both negative and positive values. Concerning those related to the business class, the most frequent patterns occur both in the weekend and during the business week, but more interestingly they occur with a short time interval in between during the same day of the week (i.e. Friday). Moreover, the error is mainly focused on positive value in the small positive range. Therefore, something strange happens on Friday the CIA does not understand. Jason Bourne could go investigate that!

3.2 Discussion and main contributions

The whole Visual Analytic framework is shown in Figure 6. The main advantage of such an approach is the central role of the user. S/he is directly involved in the analysis and via visual and intuitive metaphors can adjust the input parameters, as well as be able to manipulate any visual representation. As a matter of fact, the bubbles (events), the rings (intervals) and the connecting segments (which represent multi-time interval patterns), can be moved e.g., to obtain a better display, or to focus on some patterns. Moreover, the interactions with "day of the week" and "day of the month"[10], as well as the possibility to have detailed information (via pattern list instances) allow to gain more insights from the available data. In this way, the user may perform an explorative analysis, but at the same time can confirm some hypothesis s/he might have built apriori.

The second main contribution concerns the intertwined used of analytical and visual methods. In the framework, visual and analytical abilities are combined in order not only to present the results of the analysis, but also and rather to support the user during the whole analysis process. In fact, the framework provides intuitive and interactive ways to conduct an analysis, to proceed back and forth throughout the process, to adjust the parameters and obtain a dynamical response from all the active windows in the framework.

However, the framework still presents some limitations. First of all, one or two visualizations would help to represent quantitative information which is now represented in a tabular form. Moreover, the adoption of further visualizations would also help to provide other views of the available data. Furthermore, all active visualizations would be connected one another via brushing and linking. In this way, any interaction or selection of visual features can be overall effective and the framework is able to provide continually a coherent view of the data.

4. Conclusions

In this short paper we presented a visual analytics framework and its prototypical implementation. Though in work-in progress and with some parts lacking, this work clearly outlines the advantages coming from the collaboration of visual and analytical methods. Moreover, it encompasses the importance of the user when interacting directly with the visual representation and the role of interaction itself when dealing with time-oriented data. Furthermore, even if Ludlum's hero could not apply Visual Analytical methods to catch the Jackal, we demonstrated that such abilities would have been of help to this aim.

Finally, next steps of our work will be devoted to the improvement of the framework itself, as well as to the addition of new visual and possibly analytical methods. A usability study will be conducted in order to obtain real users' responses. Starting from such results we could then improve the capabilities and the features of the framework.

Acknowledgement

This work was supported by the program "FIT-IT Visual Computing" (Project number: 813388) as well as by the EU VisMaster Coordination Action in the FET programme (Grant number 225429).

References

[TC05] THOMAS J. and COOK K. Illuminating the path: The research and development agenda for visual analytics. IEEE, 2005.

[RFP09] RIBARSKY W., FISHER B., and POTTENGER W. M.. Science of Analytical Reasoning. Information Visualization, 8:254–262, 2009.

[BL09] BERTINI E. and LALANNE D. Surveying the Complementary Roles of Automatic Data Analysis and Visualization in Knowledge Discovery, Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge (VAKD09) Discovery, Paris, France, Pp. 12–20, June 28, 2009.

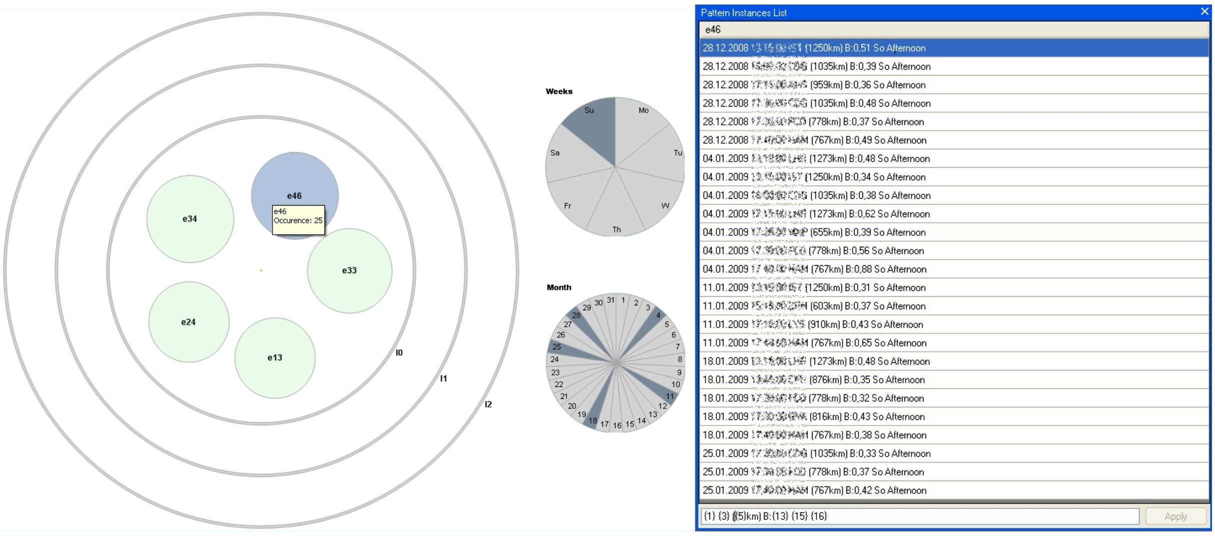
[LKL*04] LIN, J., KEOGH, E., LONARDI, S., LANKFORD, J. P. and NYSTROM, D. M. (2004). Visually Mining and Monitoring Massive Time Series. In proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle, WA, Aug 22-25, 2004.

[CCT03] CHITTARO L., COMBI C., TRAPASSO G., Data Mining on Temporal Data: a visual Approach and its Clinical Application to Hemodialysis, Journal of Visual Languages and Computing, vol.14, no.6, pp.591-620, December 2003.

[SS05] SEO J. and SHNEIDERMAN B., "Knowledge Integration Framework for Information Visualization," LNCS, Vol. 3379, pp. 207-220, Springer-Verlag, Berlin Heidelberg New York, 2005.

[Ber09] BERTONE, A.: A Matter of Time: Multi-time Interval Pattern Discovery to Preserve the Temporal Information in between, Supervisors: Silvia Miksch (Danube University Krems), Margit Pohl (Vienna University of Technology), December, 2009.

[10] We focussed on "day of the week" and "day of the month" for our analyses, but according to the considered granularities, other representations could be added, e.g., month of the year, along with their visual counterparts.



e46

28.12.2008 [redacted] (1250km) B:0,51 So Afternoon
28.12.2008 [redacted] (1035km) B:0,39 So Afternoon
28.12.2008 [redacted] (959km) B:0,36 So Afternoon
28.12.2008 [redacted] (1035km) B:0,48 So Afternoon
28.12.2008 [redacted] (778km) B:0,37 So Afternoon
28.12.2008 [redacted] (767km) B:0,49 So Afternoon
04.01.2009 [redacted] (1273km) B:0,48 So Afternoon
04.01.2009 [redacted] (1250km) B:0,34 So Afternoon
04.01.2009 [redacted] (1035km) B:0,38 So Afternoon
04.01.2009 [redacted] (1273km) B:0,62 So Afternoon
04.01.2009 [redacted] (655km) B:0,39 So Afternoon
04.01.2009 [redacted] (778km) B:0,56 So Afternoon
04.01.2009 [redacted] (767km) B:0,88 So Afternoon
11.01.2009 [redacted] (1250km) B:0,31 So Afternoon
11.01.2009 [redacted] (603km) B:0,37 So Afternoon
11.01.2009 [redacted] (910km) B:0,43 So Afternoon
11.01.2009 [redacted] (767km) B:0,65 So Afternoon
18.01.2009 [redacted] (1273km) B:0,48 So Afternoon
18.01.2009 [redacted] (876km) B:0,35 So Afternoon
18.01.2009 [redacted] (778km) B:0,32 So Afternoon
18.01.2009 [redacted] (816km) B:0,43 So Afternoon
18.01.2009 [redacted] (767km) B:0,38 So Afternoon
25.01.2009 [redacted] (1035km) B:0,33 So Afternoon
25.01.2009 [redacted] (778km) B:0,37 So Afternoon
25.01.2009 [redacted] (767km) B:0,42 So Afternoon

{1} {3} ({5}km) B:{13} {15} {16}

Figure 3: *On the left: Events above the given threshold are shown (Step 0). The size of the bubbles represents the number of occurrence of each event. The selection of an event, e.g., e_{46}, highlights the graphical representation of Day of the Week and Day of the Month. In this case, e_{46} occurs on Sunday and on the 4^{th}, 11th, 18th, 25th, 28th day of the month (note that according to the considered granularities, other representations could be added, e.g., month of the year). On the right: The Event/Pattern Instances List window shows the instances of the selected event or pattern (Note that the data in the figure have been anonymized due to privacy reasons).*

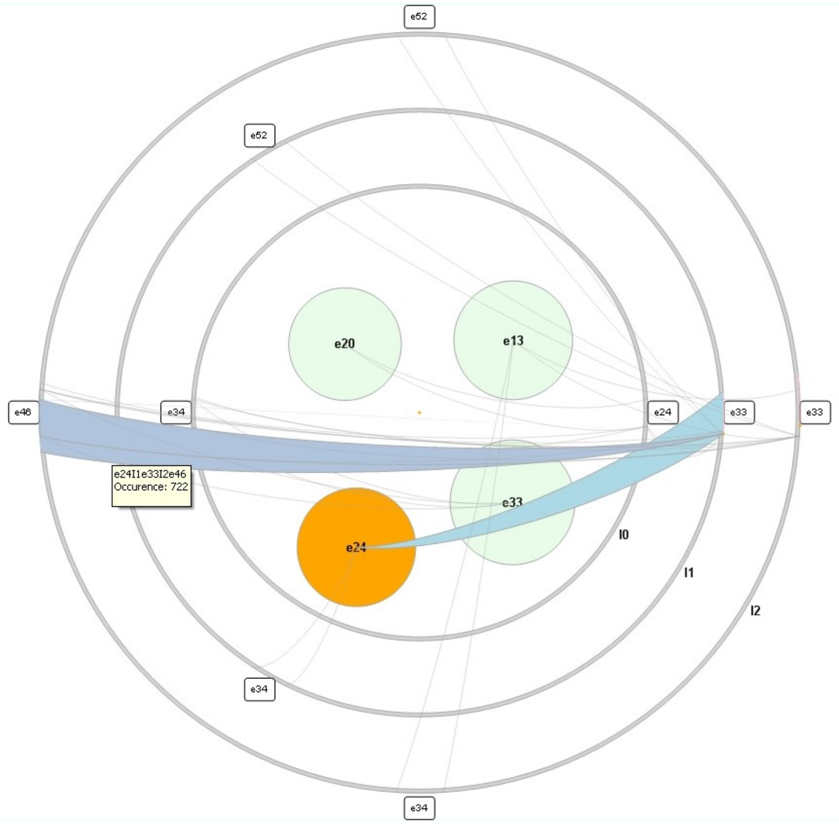
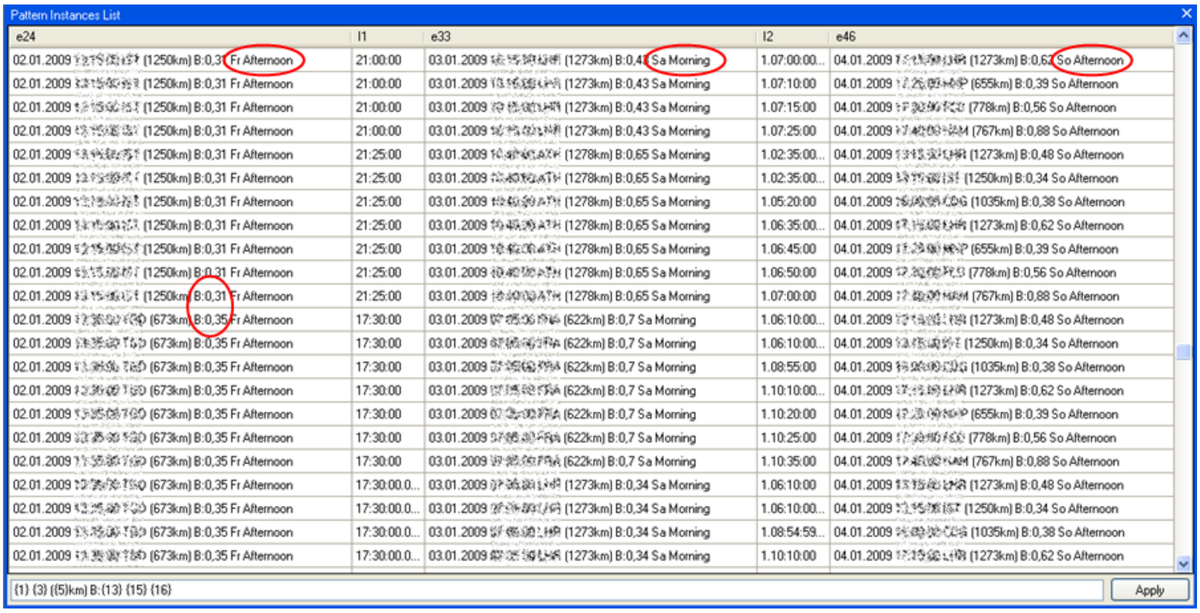


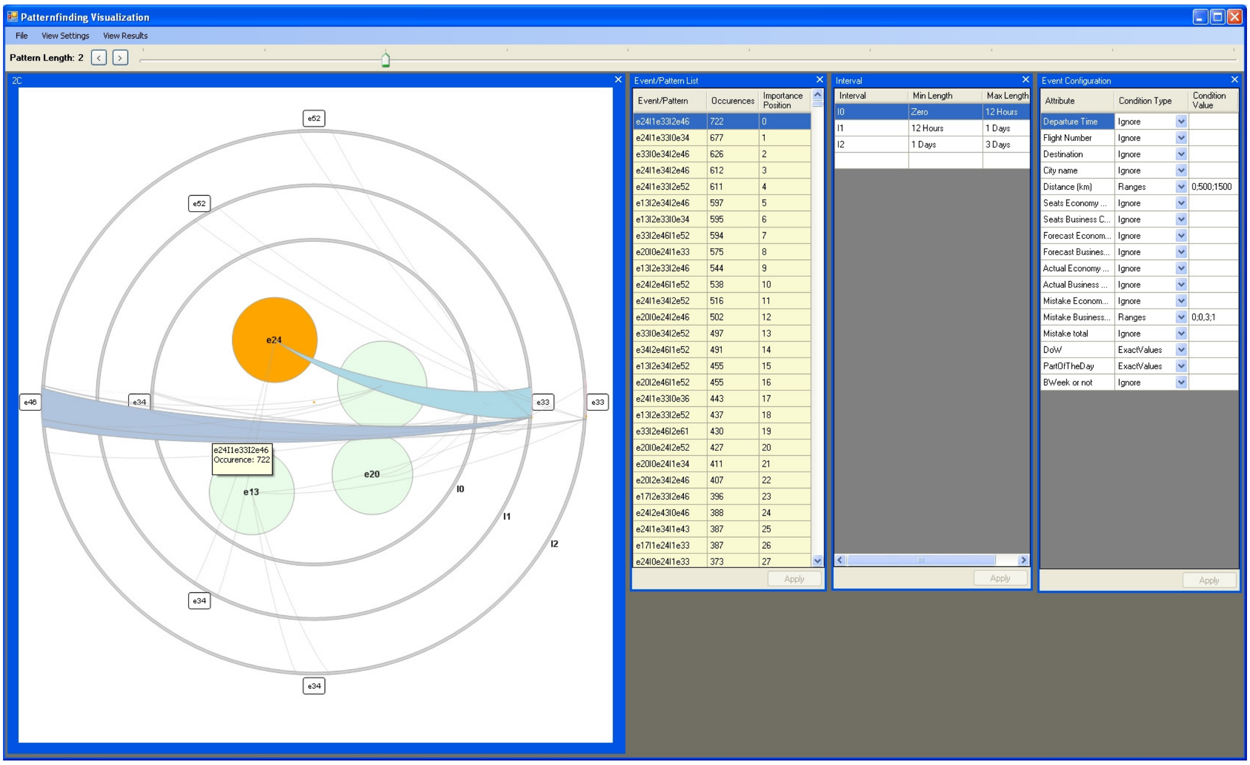
Figure 4: *Multi-time interval patterns of length 2.*



e24	I1	e33	I2	e46
02.01.2009 [redacted] (1250km) B:0,31 Fr Afternoon	21:00:00	03.01.2009 [redacted] (1273km) B:0,43 Sa Morning	1.07:00:00...	04.01.2009 [redacted] (1273km) B:0,62 So Afternoon
02.01.2009 [redacted] (1250km) B:0,31 Fr Afternoon	21:00:00	03.01.2009 [redacted] (1273km) B:0,43 Sa Morning	1.07:10:00	04.01.2009 [redacted] (655km) B:0,39 So Afternoon
02.01.2009 [redacted] (1250km) B:0,31 Fr Afternoon	21:00:00	03.01.2009 [redacted] (1273km) B:0,43 Sa Morning	1.07:15:00	04.01.2009 [redacted] (778km) B:0,56 So Afternoon
02.01.2009 [redacted] (1250km) B:0,31 Fr Afternoon	21:00:00	03.01.2009 [redacted] (1273km) B:0,43 Sa Morning	1.07:25:00	04.01.2009 [redacted] (767km) B:0,88 So Afternoon
02.01.2009 [redacted] (1250km) B:0,31 Fr Afternoon	21:25:00	03.01.2009 [redacted] (1278km) B:0,65 Sa Morning	1.02:35:00...	04.01.2009 [redacted] (1273km) B:0,48 So Afternoon
02.01.2009 [redacted] (1250km) B:0,31 Fr Afternoon	21:25:00	03.01.2009 [redacted] (1278km) B:0,65 Sa Morning	1.02:35:00...	04.01.2009 [redacted] (1250km) B:0,34 So Afternoon
02.01.2009 [redacted] (1250km) B:0,31 Fr Afternoon	21:25:00	03.01.2009 [redacted] (1278km) B:0,65 Sa Morning	1.05:20:00	04.01.2009 [redacted] (1035km) B:0,38 So Afternoon
02.01.2009 [redacted] (1250km) B:0,31 Fr Afternoon	21:25:00	03.01.2009 [redacted] (1278km) B:0,65 Sa Morning	1.06:35:00...	04.01.2009 [redacted] (1273km) B:0,62 So Afternoon
02.01.2009 [redacted] (1250km) B:0,31 Fr Afternoon	21:25:00	03.01.2009 [redacted] (1278km) B:0,65 Sa Morning	1.06:45:00	04.01.2009 [redacted] (655km) B:0,39 So Afternoon
02.01.2009 [redacted] (1250km) B:0,31 Fr Afternoon	21:25:00	03.01.2009 [redacted] (1278km) B:0,65 Sa Morning	1.06:50:00	04.01.2009 [redacted] (778km) B:0,56 So Afternoon
02.01.2009 [redacted] (1250km) B:0,31 Fr Afternoon	21:25:00	03.01.2009 [redacted] (1278km) B:0,65 Sa Morning	1.07:00:00	04.01.2009 [redacted] (767km) B:0,88 So Afternoon
02.01.2009 [redacted] (673km) B:0,35 Fr Afternoon	17:30:00	03.01.2009 [redacted] (622km) B:0,7 Sa Morning	1.06:10:00...	04.01.2009 [redacted] (1273km) B:0,48 So Afternoon
02.01.2009 [redacted] (673km) B:0,35 Fr Afternoon	17:30:00	03.01.2009 [redacted] (622km) B:0,7 Sa Morning	1.06:10:00...	04.01.2009 [redacted] (1250km) B:0,34 So Afternoon
02.01.2009 [redacted] (673km) B:0,35 Fr Afternoon	17:30:00	03.01.2009 [redacted] (622km) B:0,7 Sa Morning	1.08:55:00	04.01.2009 [redacted] (1035km) B:0,38 So Afternoon
02.01.2009 [redacted] (673km) B:0,35 Fr Afternoon	17:30:00	03.01.2009 [redacted] (622km) B:0,7 Sa Morning	1.10:10:00...	04.01.2009 [redacted] (1273km) B:0,62 So Afternoon
02.01.2009 [redacted] (673km) B:0,35 Fr Afternoon	17:30:00	03.01.2009 [redacted] (622km) B:0,7 Sa Morning	1.10:20:00	04.01.2009 [redacted] (655km) B:0,39 So Afternoon
02.01.2009 [redacted] (673km) B:0,35 Fr Afternoon	17:30:00	03.01.2009 [redacted] (622km) B:0,7 Sa Morning	1.10:25:00	04.01.2009 [redacted] (778km) B:0,56 So Afternoon
02.01.2009 [redacted] (673km) B:0,35 Fr Afternoon	17:30:00	03.01.2009 [redacted] (622km) B:0,7 Sa Morning	1.10:35:00	04.01.2009 [redacted] (767km) B:0,88 So Afternoon
02.01.2009 [redacted] (673km) B:0,35 Fr Afternoon	17:30:00.0...	03.01.2009 [redacted] (1273km) B:0,34 Sa Morning	1.06:10:00	04.01.2009 [redacted] (1273km) B:0,48 So Afternoon
02.01.2009 [redacted] (673km) B:0,35 Fr Afternoon	17:30:00.0...	03.01.2009 [redacted] (1273km) B:0,34 Sa Morning	1.06:10:00...	04.01.2009 [redacted] (1250km) B:0,34 So Afternoon
02.01.2009 [redacted] (673km) B:0,35 Fr Afternoon	17:30:00.0...	03.01.2009 [redacted] (1273km) B:0,34 Sa Morning	1.08:54:59...	04.01.2009 [redacted] (1035km) B:0,38 So Afternoon
02.01.2009 [redacted] (673km) B:0,35 Fr Afternoon	17:30:00.0...	03.01.2009 [redacted] (1273km) B:0,34 Sa Morning	1.10:10:00	04.01.2009 [redacted] (1273km) B:0,62 So Afternoon

{1} {3} ({5}km) B:{13} {15} {16}

Figure 5: *The Event/Pattern Instances List window shows that long weekend patterns and an error greater than 0.3 are clearly outlined (Note that the data in the figure have been anonymized due to privacy reasons).*



Pattern Length: 2

Event/Pattern	Occurences	Importance Position
e24I1e33I2e46	722	0
e24I1e33I0e34	677	1
e33I0e34I2e46	626	2
e24I1e34I2e46	612	3
e24I1e33I2e52	611	4
e13I2e34I2e46	597	5
e13I2e33I0e34	595	6
e33I2e46I1e52	594	7
e20I0e24I1e33	575	8
e13I2e33I2e46	544	9
e24I2e46I1e52	538	10
e24I1e34I2e52	516	11
e20I0e24I2e46	502	12
e33I0e34I2e52	497	13
e34I2e46I1e52	491	14
e13I2e34I2e52	455	15
e20I2e46I1e52	455	16
e24I1e33I0e36	443	17
e13I2e33I2e52	437	18
e33I2e46I2e61	430	19
e20I0e24I2e52	427	20
e20I0e24I1e34	411	21
e20I2e34I2e46	407	22
e17I2e33I2e46	396	23
e24I2e43I0e46	388	24
e24I1e34I1e43	387	25
e17I1e24I1e33	387	26
e24I0e24I1e33	373	27

Interval	Min Length	Max Length
I0	Zero	12 Hours
I1	12 Hours	1 Days
I2	1 Days	3 Days

Attribute	Condition Type	Condition Value
Departure Time	Ignore	
Flight Number	Ignore	
Destination	Ignore	
City name	Ignore	
Distance (km)	Ranges	0;500;1500
Seats Economy ...	Ignore	
Seats Business C...	Ignore	
Forecast Econom...	Ignore	
Forecast Busines...	Ignore	
Actual Economy ...	Ignore	
Actual Business ...	Ignore	
Mistake Econom...	Ignore	
Mistake Business...	Ranges	0;0,3;1
Mistake total	Ignore	
DoW	ExactValues	
PartOfTheDay	ExactValues	
BWeek or not	Ignore	

Figure 6: *The whole framework as it may appear during Step 2: (from left to right) 2C visualization for the representation of multi-time interval patterns; the Event/Pattern List window shows the found multi-time interval patterns; the Interval window lets the user define the intervals; the Event Configuration window which allows to choose the attributes of interest in order to define the events and possibly to discretize one or more of them (e.g., Distance, Mistake Business). The bar on the top allows to move back and forth from a step to another one. All the changes in the configuration can be dynamically applied in an interactive manner.*