Proceedings of the 3rd Workshop on Intelligent Techniques for Web Personalization (ITWP'05), in conjunction with the 19th International Joint Conference on Artificial Intelligence (IJCAI'05), Edinburgh, Scotland, UK. August 1, 2005. (Bamshad Mobasher and Sarabjot Singh Anand, editors).

Contextual Comparison of Discovered Web Mining Patterns

A.G. Büchner, M. Baumgarten, W.D. Patterson

Northern Ireland Knowledge Engineering Laboratory
European Centre of Excellence for Web Mining
Faculty of Engineering, University of Ulster
{ag.buchner, m.baumgarten, wd.patterson}@ulster.ac.uk

Abstract

The paper proposes a novel method for the discovery of interesting results in web mining applications. Contextual comparison of discovered web mining patterns deals with the interpretation of outputs from knowledge discovery algorithms applied to online data. The idea is to define the *context* of individual results based on quantifiable application-specific phenomenon of interest, associate an interestingness value to *changes* of the values this phenomenon can take, and then observe how the interestingness changes when observing the result from different contexts (such as time periods, parameter settings or data samples) and for different contexts (such as biases and user viewpoints).

1 Introduction

Dissemination, application and deployment of results generated by web mining exercises, have been of considerable interest to the research community. The main objective is that once useful and novel information has been discovered from web-based information systems, it is utilized for analysis, recommendation, customization and personalization. A problem which occurs frequently in both, commercial and research scenarios, is that results from web mining exercises carried out or analyzed in separate circumstances have to be compared. The objective of this paper is to address this issue.

In order to illustrate the problem, imagine the following scenario. A detailed analysis is carried out on the visitor base of a company's web site for the month of January. Based on the findings, a number of structural changes are implemented, which see some pages amalgamated, some pages removed and other pages extended with more content and links to related pages. After the introduction of the recommended modifications, another analysis is carried out in February. The question one would ask subsequently is "How much better are the results based on the introduced changes?" However, no mechanism exists at present to carry out such a comparison operation. Also, "better" in the view of one department might not necessarily mirror the view of another division. The concept of different contexts is able to shed light behind those queries. Similarly, when evaluating different algorithms for the discovery of the same type of patterns or the same algorithm set to different parameters, the results cannot be compared without mundanely stepping through the details of each result set.

Section 2 defines the problem and outlines the scope of the paper. In Section 3, related work is recapitulated and drawbacks are shown. In Section 4, a generic contextual interestingness framework is presented and notational issues are addressed. Section 5 applies the proposed framework to a range of knowledge types including web mining specific navigational measures. An application in the web usage mining arena is presented in Section 6. Section 7 concludes the paper and outlines future work.

2 Problem Definition

Results from knowledge discovery in general and web mining in particular are generated or interpreted in different contexts. The types of context which are relevant are *algorithm contexts* (same data is run through different algorithms, for example ID3 and C5), *data contexts* (different data is used with the same algorithm and identical threshold settings, for instance from different time spans or samples), *parameter contexts* (parameters such as thresholds are modified using the same algorithm with the same data), or any permutation thereof. Additionally, analysts will interpret the results from different viewpoints (*user context*).

In order to allow the interpretation of results generated in such disparate situations, it is necessary to have access to a flexible, yet powerful, mechanism which allows the comparison of knowledge generated by different types of web mining exercises. Issues which arise are

- What types of knowledge can be compared with each other?
- How can contextual information be incorporated by the user?
- What is the most appropriate equivalence mechanism to be applied in order to perform comparisons?

This work resolves these issues and provides a contextual interestingness measure which can be used for comparison of results from knowledge discovery. This mechanism is then extended with web mining specific concepts to allow the comparison of knowledge generated from online data.

The high-level calculation of contextual interestingness θ of any knowledge component of a certain *type* in a certain *context* is formulated as follows.

$$\theta_{Type}^{Context} \rightarrow [0..1] \qquad (1)$$

The greater the value of θ, the more interesting it is. The objective is to specify this calculation with the greatest degree of flexibility and support for context.

3 Related Work

[Silberschatz and Tuzhilin, 1996] have tackled the central problem of 'good' measures to identify the interestingness of a pattern. Two different kinds of interestingness have been introduced. Objective measurements relate to the structure of a pattern object and the underlying data used to discover them, while subjective measurements depend on the user's needs, the domain the data is analyzed in, and the scenario to which they are applied. While the approach allows the comparison of interestingness values, it neither provides a vehicle to define the concept of comparative or contextual interestingness, nor is it applicable to most web mining applications. Further related work can be sub-divided into three main groups.

Knowledge fusion (and ensemble methods) deals with the combination of knowledge. Different approaches have been presented for different knowledge types, for instance classification [Kittler and Roli, 2002], neural networks [Sharkey, 1999], and cluster ensembles [Stehl and Ghosh, 2002]. However, knowledge fusion allows the virtual merging of knowledge per se, while the approach presented in this paper uses the results of an applied model.

Knowledge sharing refers to the process of locating and extracting knowledge from multiple, heterogeneous sources, and transforming it so that the union of the knowledge can be applied in problem-solving. A good overview of different enabling technologies is given in [Neches *et al.*, 1991]. The most common approaches include mediation [Wiederhold, 1992], agent-based collaboration [Finin *et al.*, 1994], and ontologies [Gruber, 1995], or any combination thereof. Similarly, this discipline is operates on the knowledge level, which allows the virtual combination of knowledge from different resources, while the approach presented in this paper uses the actual results as basis for comparison.

Sequence Alignment Methods (SAM) have been applied in a similar context as our approach. The objective of SAM is to calculate the (string edit) distance or similarity between two sequences, which is reflected by the numbers necessary to convert a source sequence into its target counterpart. SAM has been applied to clusters [Wang and Zaïane, 2002] and sequences [Hay *et al.*, 2003]. While the approach allows the comparison of knowledge results over time, it is limited by three accounts. Firstly, the defined distance measure does not reflect the difference between two knowledge sets; it only provides an artificial number. Secondly, the method is static in that it does not allow any user intervention, which limits it for knowledge comparison applications. Thirdly, the method is limited to certain knowledge types, namely sequencing and segmentation.

4 Contextual Interestingness Framework

This section presents the framework which provides structures and operations for the comparison of multiple results from knowledge discovery [Büchner *et al.*, 2004]. *The principle idea is not to compare knowledge per se, but to compare the results which are derived from discovered knowledge.* This concept is key to the comparison mechanism presented in here. While approaches have been presented which compare knowledge as such (see Section 3), they are perceived to be too restrictive for web mining and personalization applications. They either require a high degree of compatibility between entities of knowledge or restrict the comparison on a highly granular level.

4.1 Result Comparison Structure

The outcome of a knowledge discovery exercise is a model that is of either predictive or descriptive nature. Furthermore, each model is defined to be of a certain type and format, for instance a neural network or a set of sequences. While models of some type contain information about the data they have been derived from (associations, sequences, episodes), most types only provide information about the model itself (rules, clusters, neural networks, regression, and so on). Although the formats of different model types are incompatible with each other, they all follow the same basic output philosophy, which is represented by pattern objects and descriptive attributes acting as meta-knowledge.

While it is in principle possible to compare results of different types for example, a neural network with a decision tree, the scope of this work is restricted to the comparison of compatible results, that is, results of the same type. All results of the same type t are organized in a result space $\Re$.

Definition 1. Result Space $\Re$

$\Re = \{R_1, R_2, R_3, \ldots\}$, such that $\left|\bigcup_{i=1}^{|\Re|} t(R_i)\right| = 1$. ♦

Each result $R \in \Re$ contains a set of result elements which describe the result with quantitative and / or qualitative values. Quantitative measures represent information about the result element per se, such as support, confidence and number of visits. Qualitative measures provide information about the content of result elements, for example the quantification of values through weights or costs.

Definition 2. Result R and Result Elements r

$R = \{r_1, r_2, r_3, \ldots\}$; each result element $r = \langle I, a\rangle$, where I is an optional set of items and a a set of attribute tuples such that $a = \{\langle\lambda_1, \upsilon_1\rangle, \langle\lambda_2, \upsilon_2\rangle, \ldots\}$, where λ represents a label and υ its normalized value $(0 \le \upsilon \le 1)$*. ♦

Quantitative and qualitative values are treated holistically and are referred to as attributes. A set of attributes can be attached to each element in the result space hierarchy. The range of available attributes depends on the type of knowledge that has been generated. A topology containing results and allotted attributes is depicted in Figure 1 below.

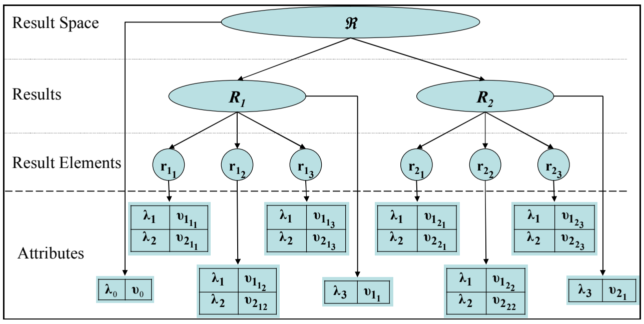

Figure 1. Result Space Topology

While the result elements layer is the lowest level within the topology it does not have to be the lowest level with respect to the available knowledge. Each *r* corresponds to a single entity of knowledge that can be of primitive or complex structure. For instance, a pattern can contain a single number representing a metric or a complex object such as a multi-dimensional sequence, embracing any number of itemsets at different dimensions, thus forming additional levels within the result elements layer. While the set of attributes can vary for different levels, it is compulsory that elements that are on the same level have the same set of attributes, which is guaranteed through the fact that each *R* has to be of the same type.

The result space topology is the cornerstone of the framework in which all elements are represented in a transparent and model type agnostic way, which allows the flexible interaction with users.

In order to illustrate the outlined concepts two examples are provided dealing with the results of neural networks and decision trees, respectively. The data originates from a newly introduced e-learning course in which the drop-out rate is higher that in its face-to-face equivalent. The objective of both exercises is to apply a model on two data sets (Q1'Y1 and Q1'Y2) of the same students, which predicts the likelihood of a student abandoning the course. The input data contains information on students' time spent online, interim results of answered questions, number of viewed learning objects, overall progress, frequency of visits, age, gender, and experience using virtual learning environments.

Example 1. Neural Network Result
The output layer of the neural network provides the weight of each prediction (r_1 to r_4). The two results shown in the table below contain attributes in the form of weights. The results do not contain any items *I*; true and false values have been added for clarity (with a cut off point at 0.5).

R_1	I_{Year1}	a_{weight}
r_{11}	F	0.3
r_{12}	T	0.7
r_{13}	F	0.2
r_{14}	F	0.4

R_2	I_{Year2}	a_{weight}
r_{21}	F	0.4
r_{22}	T	0.6
r_{23}	T	0.8
r_{24}	T	0.6

Table 1. Neural Network Results

Example 2. Rule Induction Result
The rule induction counterpart has been applied with three classification labels (low, medium and high) for the predictability of a drop-out, which are represented as items. The output sets contain two attributes, namely support and confidence.

R_1	I_{Year1}	$a_{support}$	$a_{confidence}$
r_{11}	Medium	0.04	0.4
r_{12}	High	0.05	0.2
r_{13}	Low	0.07	0.8
r_{14}	Medium	0.01	0.4

R_2	I_{Year2}	$a_{support}$	$a_{confidence}$
r_{21}	High	0.05	0.5
r_{22}	High	0.02	0.8
r_{23}	Medium	0.07	0.6
r_{24}	High	0.02	0.2

Table 2. Rule Induction Results

The sample result elements of each result are shown in the two tables above. Questions that are feasible to ask are: "Has the retention rate of the course improved or deteriorated?" and "Which students' likelihood to drop out has improved / deteriorated over the last 2 years?"

4.2 Contextual Interestingness

In order to compare attributes and results, the notion of contextual interestingness is introduced at attribute, element and result level. In order to allow the user to specify contextuality for a given problem, the concept of contexts is introduced. A context describes a given phenomenon, scenario or problem, using knowledge discovery specific attributes. Contexts are organized in a context space Γ.

Definition 3. Contexts
$\Gamma = \{\gamma_1, \gamma_2, \gamma_3, \ldots\}$, where each $\gamma = \{\alpha_1, \alpha_2, \alpha_3, \ldots\}$. Each attribute $\alpha = (\lambda, \delta, \iota)$, where $\delta \in \{0, 1\}$ and $0 \leq \iota \leq 1$. ♦

The label λ is the name of a phenomenon in a given problem space, also known as context identifier. Examples are thresholds (support, confidence and recency) or quantitative information (weight, size and duration). The label is the logical link between result attributes and context attributes. The direction δ is a binary value that states whether an increase of the value is positive (0) or negative (1) in the context the comparison is carried out. The importance factor ι states the relevance of the attribute.

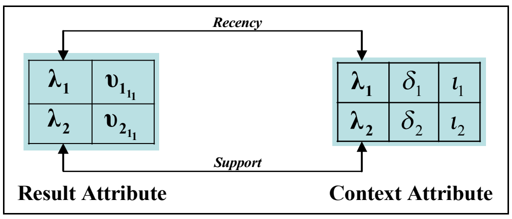

Figure 2. Result and Context Attributes Relationship

* Normalization of values is not described in here for reasons of brevity. Depending on the nature of the data, lower and upper borders can be derived automatically from data values or they can be adjusted manually. Furthermore, outliers can be treated in different ways, depending on the circumstances of the analysis.

For example, when discovering sequences in a web mining application, long sequences are attractive when the host's remuneration is based on the number of page impressions. Contrarily, short sequences are appealing if the objective is that customers solve their problem with as few clicks as possible. Both importance and direction are adjustable within the given limits.

The interestingness of an attribute comprises the degree of interest associated with it in a given context. That is, when putting a certain result in a certain context, the interestingness of each individual attribute can be calculated using the definition below.

Definition 4. Attribute interestingness θ_a

$\theta_a(\upsilon, \delta, \iota) = |(\upsilon - \delta) * \iota| \rightarrow [0\ldots1]$ ♦

Interestingness embraces both objective and subjective measurements. The former relates to the structure of a pattern and the underlying data used to discover it, while the latter depends on the user's needs, the domain the data analyzed is in, and the scenario to which it is applied. Thus, the interestingness of a pattern is by no means an objective value that remains constant across comparisons; interestingness is a subjective representation of the user's priorities in conjunction with the raw pattern values.

In order to compute the interestingness θ_o of the attributes of either a result element r or a result R all attributes are taken into account.

Definition 5. Object interestingness θ_o

$$\theta_o(a) = \left(\sum_{i=1}^{|a|} \theta_a(\upsilon_i, \delta_i, \iota_i) \right) / |a| \rightarrow [0..1]$$ ♦

$|a|$ represents the amount of attributes in a. This operation calculates the arithmetic mean of all attribute values in a given context Γ. This measure is used as basis for the calculation of the result element interestingness θ_e and the result interestingness θ_r.

Definition 6. Element interestingness θ_e

$\theta e(r) = \theta(a(r)) \rightarrow [0..1]$ ♦

Definition 7. Result interestingness θ_r

$\theta_r(R) = \theta(a(R)) \rightarrow [0..1]$ ♦

In the context of each comparison exercise, the interestingness of a knowledge component represents the appeal it holds for a given context. Thus, the resulting interestingness is an amalgamation of the original structural data, the measures derived from other levels of the topology and the context specifics dictated by the scenario of the mining exercise or the beliefs of the user. It is then obvious that the derived interestingness is a subjective measure, which can never claim to be absolute or unbiased. However, being unbiased has never been the intention, in fact bias is essential, because it dictates how comparison between components is performed and thus influences the result. Therefore, comparing knowledge components is *comparing their interestingness*. Due to the generic design of the framework, such a comparison can be applied to any two or more components on the results or result elements level. This not only lets the user decide how objects are compared with each other but also which objects one wants to compare. For instance, the set of patterns used for comparison can be limited to patterns that contain specific items or patterns that have the same start / end item, et cetera. However, it must be understood that only result components at the same level can be compared with each other. It makes no sense to compare a single sequence to a set of sequences, that is, comparing a component at result elements level with a component at results level. Only the comparison of components that possess an equivalent set of comparable measurements is reasonable.

Example 3. Contextual Comparison Calculation

Using the earlier introduced rule of the rule induction e-learning prediction example, the classification labels have been quantified to low=0.3, medium=0.6 and high=1. The following can be calculated, given that for λ_{risk} δ=0 (lower is more interesting in this context), ι=100% (highest importance), for $\lambda_{support}$ δ=1, ι=50% and for $\lambda_{confidence}$ δ=1, ι=80%.

$$\theta_e(r_{1_1}) = \frac{|(0.6-1)*1.0| + |0.04*0.5| + |0.4*0.8|}{3} = 0.247$$
$$\theta_e(r_{2_1}) = \frac{0 + |0.05*0.5| + |0.5*0.8|}{3} = 0.142 \quad (2)$$

Due to the fact that interestingness of the result in Year 1 is greater than the one of Year 2, this means that student r_1's likelihood to drop out has worsened. Or, to conform to Equation (1) – the greater the value of θ, the more interesting it is – the student's likelihood to continue with the course has increased.

Calculating the four result element interestingness measures for Year 1 and building the arithmetic mean results to 0.252. Calculating the equivalent for Year 2 results in 0.18. Those two values are accepted as new attributes λ_{risk} for R_1 and R_2, respectively (δ=0 and ι=100%). Given that most students are familiar with e-learning environments, the data set of Year 2 is 'punished' via an experience attribute (δ=1, ι=20%).

R	$\left(\sum_{i=1}^{4} \theta_e(r_i)\right)/4$	δ_1	ι_1	*age*	δ_2	ι_2
Year 1	0.252	0	100%	0.5	1	20%
Year 2	0.18	0	100%	0.8	1	20%

Table 3. Result Attributes

$$\theta_r(R_1) = \frac{|0.252*1.0| + |(0.5-1)*0.2|}{2} = 0.176$$
$$\theta_r(R_2) = \frac{|0.18*1.0| + |(0.8-1)*0.2|}{2} = 0.038 \quad (3)$$

It can be shown that the overall likelihood to abandon the course has also risen; in fact the situation has worsened substantially.

4.3 Attribute Generation

As outlined previously, the set of attributes a associated with each r and R of the result space form the basis of calculating contextual interestingness. Attributes on each level can be provided through the result sets themselves, they can be specifically set by the user or they can be derived from other attributes, for example average or coverage values. Average values represent the arithmetic mean of sub values, that is the average interestingness of all $r \in R$ calculated as follows.

$$\left(\sum_{i=1}^{|R|}\theta_e(r_i(R))\right)/\,|\,R\,| \tag{4}$$

Coverage values are derived using the scope of the overall result space. For instance, given all distinct result elements of r, a coverage attribute for each R is calculated indicating the exposure of r in each R. Equally, values can be derived for R from $\Re$. While comparisons cannot be carried out at the result space level, this justifies the existence of attributes allotted to the root level.

While the concept behind these measures is generic, the individual implementation for different types of knowledge is not. That is, different knowledge types provide different measures and for different types of knowledge different measures may be extracted at different levels of the hierarchy. Nevertheless, once the measures are populated they can be used in the same way for all knowledge types.

5 Specific Knowledge Types

Within this section, specific characteristics for different knowledge types are outlined and attributes are mapped onto the generic structure of the comparison framework. For reasons of brevity, two common patterns (clusters and rules) are presented which are frequently used in personalization and web applications (such as the e-learning example) and one web specific navigation pattern is introduced. Attributes for a wide range of models can be found in [Büchner, 2004].

The set of attributes portrayed for each knowledge type is by no means complete but reflects quantitative attributes that are generically available. Qualitative attributes are useful in the context of this work, since they are domain specific. There exists a wide variety of web-related measures for pages (such as number of views, stickiness, abandonment), visits (duration, activity, page requests), visitors (recency, frequency, life-time value) as well as metrics which are industry-specific, for instance in online retail, e-learning or workflow applications. Web-specific attributes can be allotted to all knowledge types, but are not described in more detail for reasons of brevity.

5.1 Segments / Clusters

A segment-specific topology is depicted in Figure 3, where a cluster represents a result element and its elements are part of the attributes layer. Attributes for clusters include the density of a cluster, the distance between cluster elements to a cluster's centroid, the minimum spanning distance of all elements as well as support and confidence measures.

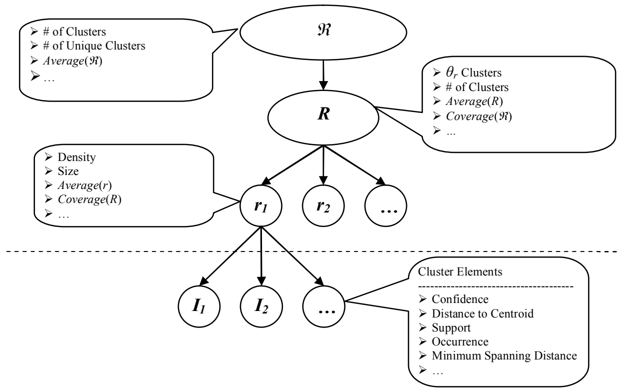

Figure 3. Segment-specific Structure

5.2 Rules

Similar to segments, each rule set resides on the result level and each rule represents a corresponding result element. The fact that each rule contains an antecedent and a consequent, which in turn contain patterns themselves, is irrelevant for the structure of the hierarchy. However, this structure can be used to aid the comparison process and to populate different attributes. In Figure 4 a rule specific topology is depicted, where a number of different attributes have been added to each level. Each result element is split into two parts containing the antecedent and the consequent, which in turn contain patterns that are used for comparison purposes.

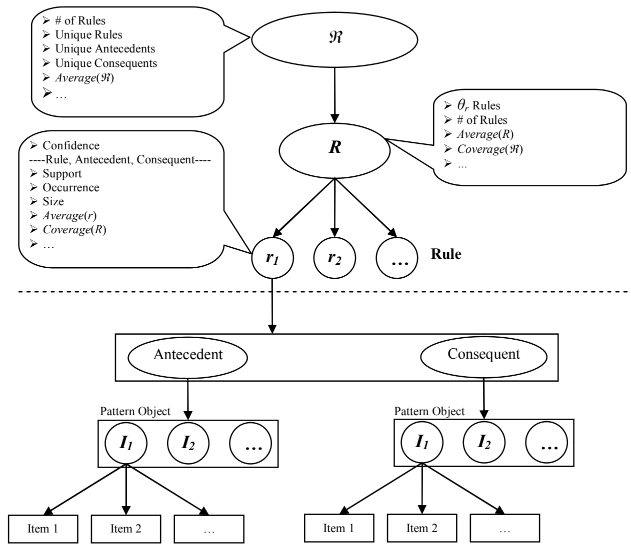

Figure 4. Rule-specific Structure

5.3 Web Navigation Patterns

Web navigation patterns can incorporate both, associative and sequential structures (multi-dimensional patterns). Attributes of both knowledge types can therefore be used at different levels of the framework. Furthermore, the general concept of such measures and the interestingness thereof can be extended into the attributes layer, which is shown in

Figure 5, introducing the concept of context at different levels of the result element itself. While this enables a more detailed specification of interestingness for individual patterns, it does not require the extension of the framework.

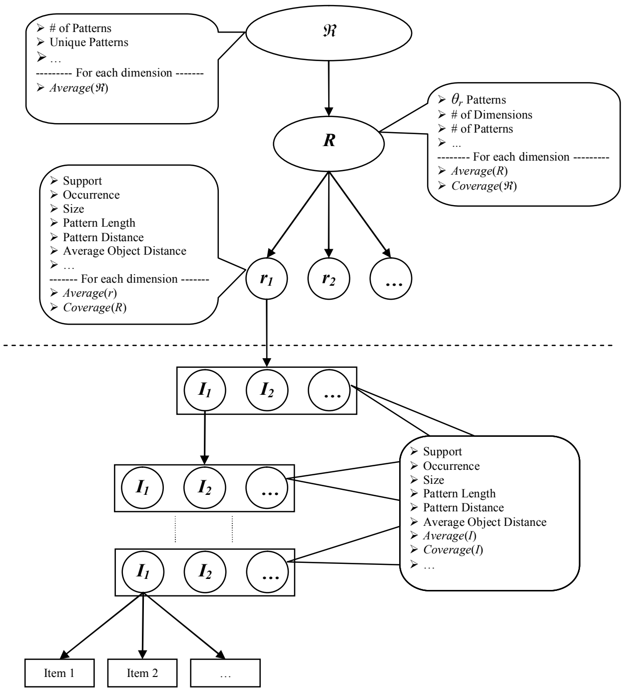

Figure 5. Web Navigation Pattern Structure

6 Application

Using the proposed framework, an application has been conducted that analyses the impact of a promotion activity over a period of three months. The problem given was to show how changes of activity relate to the promotion undertaken. Three result sets have been generated, of which a sub-set is used here (see Table 4). The first result set (R_1) represents activity prior to the promotion activity; R_2 and R_3 represent the activity after the promotion.

Each set contains a number of sequential patterns reflecting the browsing behavior over the given period of time. Patterns have been constrained to start with "Index" and to include "Pricing" at least once, thus indicating the interest in buying a product. Each pattern has an attached set of properties, namely length, support and duration indicated by ω_L, ω_S and ω_D, respectively for which non-normalized values are listed. The goals of the comparison exercises are

- to analyze the impact of the marketing campaign in general ($Context_1$),
- to identify if the general interest in the range of products has increased or not, which is mainly reflected through the number of visitors that have accessed the web site ($Context_2$), and
- to analyze if the time visitors have spend on the web site has increased or decreased ($Context_3$).

R	ID	ω_L	ω_S	ω_D	
1	1	3	31	250	⟨Index, WhitePapers, Pricing⟩
	2	7	25	160	⟨Index, Products, WhitePapers, Pricing, Evaluation, Purchase, Download⟩
	3	6	15	170	⟨Index, Company, WhitePapers, Pricing, Evaluation, Register⟩
	6	3	37	280	⟨Index, Pricing, Purchase⟩
	7	8	22	280	⟨Index, Company, WhitePapers, Pricing, Evaluation, Purchase, Download, Register⟩
	8	5	31	120	⟨Index, Products, Pricing, Download, Register⟩
	9	6	16	70	⟨Index, WhitePapers, Company, Pricing, Purchase, Download⟩
2	1	3	41	350	⟨Index, WhitePapers, Pricing⟩
	2	7	55	360	⟨Index, Products, WhitePapers, Pricing, Evaluation, Purchase, Download⟩
	3	6	65	870	⟨Index, Company, WhitePapers, Pricing, Evaluation, Register⟩
	4	7	68	180	⟨Index, Support, Index, WhitePapers, Products, Pricing, Register⟩
	5	10	62	730	⟨Index, Pricing, Index, Company, WhitePapers, Pricing, Evaluation, Purchase, Download, Register⟩
	6	3	77	880	⟨Index, Pricing, Purchase⟩
	7	8	55	280	⟨Index, Company, WhitePapers, Pricing, Evaluation, Purchase, Download, Register⟩
	8	5	41	220	⟨Index, Products, Pricing, Download, Register⟩
	9	6	70	170	⟨Index, WhitePapers, Company, Pricing, Purchase, Download⟩
	10	2	12	270	⟨Index, Pricing⟩
3	1	3	51	650	⟨Index, WhitePapers, Pricing⟩
	2	7	45	860	⟨Index, Products, WhitePapers, Pricing, Evaluation, Purchase, Download⟩
	3	6	46	770	⟨Index, Company, WhitePapers, Pricing, Evaluation, Register⟩
	4	7	38	980	⟨Index, Support, Index, WhitePapers, Pricing, Download, Register⟩
	5	10	32	630	⟨Index, Pricing, Index, Company, WhitePapers, Pricing, Evaluation, Purchase, Download, Register⟩
	6	3	57	780	⟨Index, Pricing, Purchase⟩
	7	8	35	880	⟨Index, Company, WhitePapers, Pricing, Evaluation, Purchase, Download, Register⟩
	8	5	41	520	⟨Index, Products, Pricing, Download, Register⟩
	9	6	43	570	⟨Index, WhitePapers, Company, Pricing, Purchase, Download⟩
	10	2	29	970	⟨Index, Pricing⟩

Table 4. Application Patterns

The attributes selected for comparison are shown in Table 5 together with their respective importances and directions of interest.

ID	Label	Level	$Context_1$		$Context_2$		$Context_3$	
			ι	δ	ι	δ	ι	δ
1	Number of navigational paths	*R*	100%	↓	25%	↓	25%	↓
2	Avg. number of page hits	*R*	100%	↑	25%	↑	25%	↑
3	Avg. duration	*R*	100%	↑	10%	↑	100%	↑
4	Avg. dupport	*R*	100%	↑	100%	↑	10%	↑
5	Avg. θ_r	*R*	100%	↑	100%	↑	100%	↑
6	Duration	*r*	100%	↑	10%	↑	100%	↑
7	Support	*r*	100%	↑	100%	↑	10%	↑
8	Number of page hits	*r*	100%	↓	25%	↓	25%	↓

Table 5. Application Attributes

To analyze the impact of the marketing campaign in general ($Context_1$), all measures are deemed equally important, with ι set to 100%. Furthermore, all measures keep their original direction of interest δ except measures 1 and 8. Measure 1 represents the number of navigational paths in each result set and as such reflects the diversity of visits. A large number of patterns means that a range of different paths have been explored, which indicates that visitors encountered problems locating the information they have been looking for or the product they are interested in. Therefore, the number of patterns should be as small as possible for the current comparison context. Measure 8 represents the number of page hits in each pattern, which should be small because it indicates the most direct path and therefore a good click-to-close ratio.

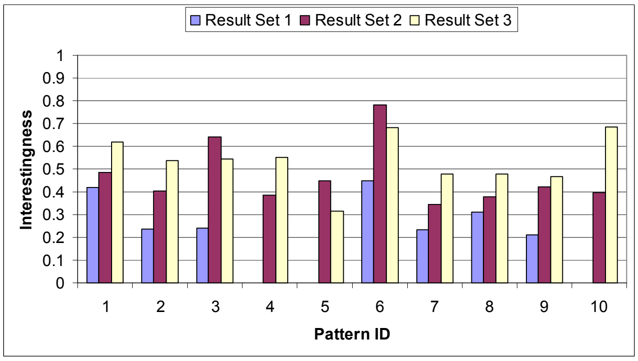

Figure 6. Comparison of Result Elements ($Context_1$)

Comparing the three result sets at element level, Figure 6, reveals how the interestingness of each pattern has changed for the current context. For all patterns, result set 1 (R_1) has the lowest interestingness. Note that patterns 4, 5 and 10 are not present in R_1 as they have not satisfied the given thresholds at the discovery phase. The comparison already shows that the marketing campaign had a positive impact and it indicates that R_3 has the highest overall interestingness because 7 of the 10 knowledge patterns are ranked highest for this result set (pattern IDs 1, 2, 4, 7, 8, 9 and 10). Comparing results at element level indicates the overall interestingness of the result sets they belong to. However, the interpretation of such a comparison exercise can be difficult if the result sets contain a large number of knowledge patterns. Furthermore, additional attributes are available at higher levels of the result space, which can be used to describe individual contexts in more detail.

$Context_1$ of Figure 7 shows the result of a comparison exercise at result level using the same parameters as above. As indicated through the previous exercise, the marketing campaign had a positive effect, which is reflected through the increase in interestingness of R_2 and R_3. However, instead of R_3, which was expected to be the most interesting set, R_2 has the highest overall interestingness value. This shows that the impact of promotion was greater in the second month than it was in the third.

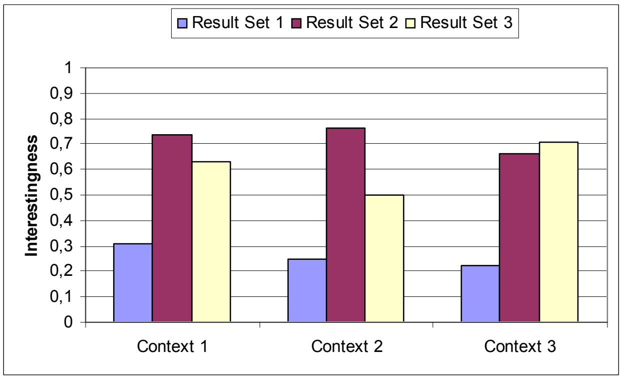

Figure 7. Comparison of Elements ($Context_{1\text{-}3}$)

The objective of the second exercise was to evaluate if the marketing campaign has increased the interest in certain products. The most influencing factor is the support attribute, which reflects the percentage of visitors that have accessed product-related pages. In order to focus the comparison, the importance of attributes 1, 2, 3, 6 and 8 have been reduced as shown in Table 5. The result of the comparison at result level is shown as $Context_2$ of Figure 7. The interest increased significantly after the introduction of the promotion but dropped slightly in the month thereafter. This behavior was expected as the effect of tailored marketing campaigns usually decreases over time.

The third exercise evaluated how the behavior of visitors changed with respect to the average visit duration. The importance of different measures has been modified accordingly, as shown in Table 5 and the result of the exercise is shown as $Context_3$ in Figure 7. The interestingness increased significantly for R_2 and increased further for R_3. Unlike in the second comparison, which indicated a drop in interest for the third month, this exercise shows that the promotion has a longer impact than anticipated.

As seen for $Context_2$ and $Context_3$, some measures are more relevant than others depending on the context the comparison is undertaken for. However, the brief example also demonstrates that less important measures should not be excluded completely as they can influence the result significantly. For instance, while R_2 is more interesting in

Context$_1$ and *Context*$_2$, R_3 is more appealing for *Context*$_3$. This shows that the comparison can and should be performed for different contexts in order to reveal the impact individual attributes have.

7 Conclusions and Future Work

A new approach has been proposed in which subjective interestingness of a pattern is determined within the context in which discovery is made. The framework is algorithm agnostic, that is it covers all common types of knowledge (*generality*). Due to its flexible structure, full *extensibility* and *re-usability* are guaranteed. The vanilla approach of the model and its calculations assure *simplicity* and *integratability*. The user can adjust all contextual values interactively (*flexibility*) which provides a solid basis for a range of domains (*applicability*). Due to the fact that all comparison procedures are applied linearly on the result sets of web discovery, *scalability* of the operations can be guaranteed. Given that results of web mining applications are usually significantly smaller than the data they are derived from, comparisons can be carried out in memory (*velocity*). An architecture based on PMML input is able to facilitate the introduced concepts:

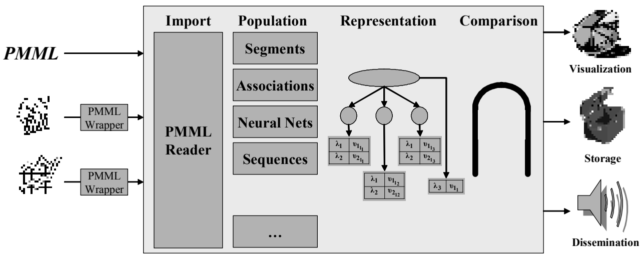

Figure 8. Comparison of Result Elements (*Context*$_1$)

The framework has been extended with web-related attributes, which allow the comparison of different types of outputs from web mining exercises. The context-driven interpretation, which can easily be automated for pre-determined conditions (contexts), provides flexible yet powerful information for intelligent personalization applications. Examples include the analysis of visitor navigations across time, detection of changes in online user behavior and the building of knowledge warehouses to be viewed from different user angles.

The handling of (un-)ambiguous dependencies between individual attributes, sets of attributes and knowledge patterns has to be addressed in the future. For instance, there is a clear dependency between support and confidence of a knowledge pattern. Moreover, a pattern such as ABC $\rightarrow$ D has a certain correlation with all its sub-patterns (AB $\rightarrow$ D, A $\rightarrow$ D, etc.). Furthermore, the dependency between parameters assigned to attributes across levels within the hierarchal result space topology has to be addressed.

Contextual interestingness measures have not only been used for comparison of results, but also for contextual ranking [Baumgarten *et al.*, 2003]. Ranking is a highly requested operation in web personalization applications. The extension of the presented framework and the amalgamation of contextual interestingness measures and context-specific attributes will provide a powerful environment in which knowledge discovered in different web contexts can be interacted with in a fully user-driven fashion.

References

[Baumgarten *et al.*, 2003] M. Baumgarten, A.G. Büchner, and J.G. Hughes. User-Driven Ranking for Measuring the Interestingness of Knowledge Patterns. *Proc. Int'l Conf. on Information and Knowledge Engineering*, pp. 122-128, 2003.

[Büchner, 2004] A.G. Büchner. Context Mediation among Knowledge Discovery Components. Universal Publishers, 2004.

[Büchner *et al.*, 2004] A.G. Büchner M. Baumgarten, J.G. Hughes, and W.D. Patterson. Contextual Comparison of Discovered Knowledge Patterns. *Proc. 16th Int'l Conf. on Software Engineering and Knowledge Engineering*, pp. 388-391, 2004.

[Finin *et al.*, 1994] T. Finin, R. Fritzson, D. McKay, and R. McEntire. Using KQML as an agent communication language. *Proc. 3rd Int'l Conf on Information and Knowledge Management*, pp. 456–463, 1994.

[Gruber, 1995] T.R. Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. *Int'l Journal of Human Computer Studies*, 43(5/6):908–928, 1995.

[Hay *et al.*, 2003] B. Hay, G. Wets, and K. Vanhoof. Segmentation of visiting patterns on websites using a Sequence Alignment Method. *Journal of Retailing and Consumer Services*, 10(3):145-153, 2003.

[Kittler and Roli, 2002] J. Kittler and F. Roli (eds.). Multiple Classifier Systems, *Lecturer Notes in Computer Science*. Vol. 2634, Springer-Verlag, 2002.

[Neches *et al.*, 1991] R. Neches, R. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator, and W.R. Swartout. Enabling Technology for Knowledge Sharing. *AI Magazine*, 12(3), Fall 1991.

[Sharkey, 1999] A. Sharkey. Combining Artificial Neural Nets. Springer Verlag, 1999.

[Silberschatz and Tuzhilin, 1996] A. Silberschatz and A. Tuzhilin. What Makes Patterns Interesting in Knowledge Discovery Systems. *IEEE Trans. on Knowledge and Data Engineering*, 8:970–974, 1996.

[Stehl and Ghosh, 2002] A. Stehl and J. Ghosh. Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions. *Journal of Machine Learning Research*, 3:583–617, 2002.

[Wang and Zaïane, 2002] W. Wang and O.R. Zaïane. Clustering Web Sessions by Sequence Alignment. *Proc. DEXA Workshop on Management of Information on the Web*, pp. 394–398, 2002.

[Wiederhold, 1992] G. Wiederhold. Mediators in the Architecture of Future Information Systems. *IEEE Computer*, 25(3):38–49, 1992.