DISCOVERY OF PERSONALIZED INFORMATION SYSTEMS USAGE PATTERNS

Inese Supulniece, Janis Grabis

Riga Technical University, Information Technology Institute, Meza 1/3, LV 1048, Riga, Latvia, Inese.Supulniece@rtu.lv, grabis@iti.rtu.lv

Abstract. Users use information systems to accomplish their tasks usually consisting of multiple steps. Each user or user group might have a preferred sequence of the steps. Adaptive information systems attempt to exploit such usage patterns, and it is expected that adaptivity can be improved through personalizing processes and accounting for temporal and sequential dependencies. However, there is a lack of empirical evidence of existence of personalized information systems usage patterns. The objective of this paper is to analyze empirical information systems usage data in order to confirm existence of personalized information systems usage patterns, which could be further used in developing adaptive information systems. Empirical data analyzed are derived from log files of customers service website of a telecommunication company and of university's e-learning system. The Longest Common Subsequence algorithm is used to discover patterns in task execution. It is found that frequency of patterns observation varies significantly between the data sets although in both cases personalized patterns are observed more frequently than general patterns. The personalized patterns also are more precise than the general patterns.

Keywords: transaction logs, temporal data mining, patterns, process adaptation.

1 Introduction

User uses information systems to perform various tasks. Each task consists of multiple steps and often there are multiple alternative sequences of steps, which can be used to accomplish the same or closely related tasks. Sequences, which are used most frequently, can be referred to as task execution patterns. Various data mining [8] and process mining [12, 13] techniques are used to discover such patterns. Data mining is able to discover hidden patterns, unexpected trends or other subtle relationships in the data using a combination of techniques from machine learning, statistics and database technologies. Temporal data mining is concerned with data mining of large sequential data sets what is important in the case of analyzing task execution sequencies (see Laxman and Sastry [8] for a recent overview of temporal data mining methods). Process mining techniques are mainly aimed at recovering process structure for process validation and performance improvement purposes.

Adaptive information systems attempt to exploit the task execution pattern to accomodate users in order to expediate execution of their tasks. The overall goal of the proposed research is to elaborate a process driven adaptive business information system, which is conceptually described in [17]. One of the main characteristics of the proposed adaptive information system is an ability to predict and to recommend future task execution steps on the basis of actions performed by the user and historically recorded behavior of this user or user groups. Some of available adaptive information systems (for example, [5, 11]) do not explore temporal and sequential dependencies and process mining techniques deal with general processes rather than user specific task execution sequences. It is expect in this research that efficiency of adaptivity can be improved through personalizing processes and accounting for temporal and sequential dependencies. However, there is a lack of empirical evidence of existence of user specific or personalized information systems usage patterns. Personalized patterns are patterns, which repeatedly occur in task execution sequences of individual users, while general patters are patters, which occur in task execution sequences regardless of users.

The objective of this paper is to analyze empirical information systems usage data in order to confirm existence of user specific information systems usage patterns, which could be further used in developing adaptive information systems. Two main hypotheses to be tested are that: 1) personalized patterns in user's task execution sequences are observed more frequently than general patterns; and 2) personalized patterns have higher level of confidence than general patterns. Two empirical information systems usage data sets are used in the paper. These data sets are derived from log files of customers service website of telecommunication company and of university's e-learning system. The Longest Common Subsequence algorithm [3] is used to discover patterns in task execution. In order to evaluate discovered patterns, two measures commonly used in data mining, namely, support and confidence [8] are used. The support measure indicates the frequency of pattern observation, and the confidence measure is a proxy for pattern precision measurement.

The rest of the paper is structured as follows. Section 2 introduces concepts and notation used in the paper. The data analysis and pattern evaluation methodology and empirical data used in the analysis are described in Section 3. Empirical data analysis is presented in Section 4. Section 5 contains related work and Section 6 concludes.

2 Information System Usage Patterns

It is assumed that a user perform various actions in the information system to accomplish her/his tasks. These actions are recorder in a log file as events. Each event is characterized by its name, time and user. Events sequentially involved within a specified timeframe form an information systems usage events sequence or session $S_i(u) \in \mathbf{S}$, where subscript *i* indicates the session number in the log file and varies from 1 to the number of sessions *N*, *u* indicates session's user and $\mathbf{S}$ is the set of all sessions in the log file. It is expected that the sessions contain information systems usage patterns. The pattern is defined as a sub-sequence of events what is present within at least two sessions. Given two sequences of events: $S_1 = \{a_1, a_2, ..., a_{T_1}\}$ and $S_2 = \{b_1, b_2, ..., b_{T_2}\}$, a pattern is such a sub-sequence $P(S_1, S_2) = \{d_1, d_2, ..., d_M\}$, where $P \subseteq S_1$ and exists integers $i_1<i_2<...<i_x$ such that $d_1 \subseteq a_{i_1}, d_2 \subseteq a_{i_2}, ..., d_M \subseteq a_{i_x}$, and $P \subseteq S_2$, where exists integers $j_1<j_2<...<j_y$ such that $d_1 \subseteq b_{j_1}, d_2 \subseteq a_{j_2}, ..., d_M \subseteq a_{j_y}$. Patterns are denoted as $P_k = \{d_{k1}, ..., d_{kM_k}\}$, where *k* is the number of pattern, d_{kj} is the *j*th event in the pattern and M_k is the number of events in the pattern. Events in the pattern might not necessarily follow each other immediately, i.e. there could be other events between them.

Patterns are classified as general patterns and personalized or user specific patterns. Let $\mathbf{S}_u$ denotes all sessions performed by user *u*, then general patterns are searched within the set $\mathbf{S} \setminus \mathbf{S}_u$, while the personalized patterns are search within the set $\mathbf{S}_u$ for each user. Personalized patterns are denoted by P_{ku}, where *k* indicates the number of the pattern and *u* indicates pattern's user.

Patterns are evaluated using their support and confidence measures. These measures are computed for each pattern separatetly. Support measures frequency of pattern occurrence within information system usage session. It is defined as

$$F_k = \frac{\sum_{i=1}^{N} Z_i}{N}, k = 1, ..., L, \quad (1)$$

where Z_i indicates presence of the *k*th pattern in the *i*th session

$$Z_i = \begin{cases} 1, \text{ if } P_k \subseteq S_i \\ 0, \text{ if } P_k \not\subset S_i \end{cases}.$$

For example, given sessions *S₁={A, B, C}, S₂={A, C, D}, S₃={D, G}*. The support for pattern *P={A, C}* is 2/3 because sequence A -> C is presented in session S_1 and S_2, but not in S_3.

Confidence is perceived as a measurement of pattern precision and is calculated

$$C_k = \frac{\sum_{i=1}^{N} Z_i}{\sum_{i=1}^{N} Y_i}, k = 1, ..., L, \quad (2)$$

where Y_i indicates presence of the right trimmed *k*th pattern $P_k^* = P_k \setminus d_{kM_k}$ in the *i*th session

$$Y_i = \begin{cases} 1, \text{ if } P_k^* \subseteq S_i \\ 0, \text{ if } P_k^* \not\subset S_i \end{cases}.$$

Essentially, the confidence measures a likelihood of a user completing the pattern once it has started to execute it. For example, given sessions $S_1=\{A, B, C\}$, $S_2=\{A, C, D\}$, $S_3=\{D, G\}$. The confidence for pattern *P={A, C}* is calculated 1 because sequence *A -> C* is presented in two session S_1 and S_2, but not in S_3 and $\sum_{i=1}^{N} Y_i$ = 2 because the pattern without the last element *P*={A}* is presented in 2 sessions, S_1 and S_2.

$F_k(u)$ and $C_k(u)$ for personalized patterns are computed using equations (1) and (2), respectively, where only sessions of the *u*th user are considered (i.e., $\mathbf{S}_u$).

3 Analysis Methodology

The aim of the analysis is to confirm presence of personalized patterns in empirical information systems usage data. The presence of such patterns would suggest that it is possible to develop adaptive information systems, where processes are adapted according to individual preferences of users rather than according to some general rules for all users or user groups. In order to confirm presence of personalized patterns, empirical data are used to test two hypotheses:

H1. Personalized patterns are observed more frequently than general patterns;

H2. Personalized patterns have higher confidence (i.e., they are more precise) than general patterns.

If both hypotheses are confirmed then individual patterns are considered more efficient than general patterns because users more frequently follow their personalized patterns and adaptation according to personalized patterns would be more accurate.

The five step methodology is developed for analysis purposes: (1) Data preparation (clean transaction logs from unnecessary information; group user activities into user sessions and transform data into format that is more convenient for analysis); (2) Finding general patterns (sequences of events that are similar to several or all users); (3) Finding individual patterns (sequences of events that are similar to each individual user); (4) Filtering of results; (5) Evaluation of general and individual patterns.

The main parts of the methodology are pattern discovery (using association rules identification algorithm) and evaluation of derived patterns. General and individual patterns are extracted from the same log files.

3.1 Data Description

Two data sources are used for analysis of information systems usage patterns, namely, log files of customers service website of telecommunication company (CS) and log files of university's e-learning system (ES).

Description of CS

CS has two functions: a representative function (anonymous user) and a service portal function (identified user). In the representative part, there is information about company, about its products, contacts and, additionally, there are a lot of advertisements (special offers) and possibilities to apply for products. In the service portal part, there are possibilities to view and pay bills, view statements, etc. There is also functionality for a webpage administrator. Thus administration tasks are also included in the log file. The portal is more a flat application compared to workflow systems, so user navigation within the portal is not restricted by some predefined flows.

The analyzed log files contain data from 3 days including user activities and also system responses. The total number of entries in the log file is ~128000 transactions (rows) and ~600 unique users (determined according to IP addresses).

Description of ES

The second set of data comes from log files of ES. This portal is used only by identified users (students and course owners). Students can review information about their classes, homeworks, grading and other teaching resources, submit their homeworks and communicate with each other. Additionally, there is also functionality for course owners to add and update course related information. Similarly as in the case of CS, there are no predefined workflows and users have high degree of flexibility.

The analyzed log files contain data from 1 month. – They include only user activities. The total number of entries in the log file is ~10000 transactions (rows) and 44 unique users.

3.2 Data Preparation

Original logs of transactions from real systems require initial cleansing to adjust formatting. Described steps are similar for any web usage mining problem and are discussed in [2]. Additionally, content of logs from real systems might include also not relevant entries. For example, the HTTP protocol requires a separate connection for every file that is requested from the Web server. Therefore, a user's request to view a particular page often results in several log entries since graphics and scripts are down-loaded in addition to the HTML file. In most cases, only the log entry of the HTML file request is relevant and should be kept for the user session file. This is because, in general, a user does not explicitly request all of the graphics that are on a Web page, they are automatically down-loaded due to the HTML tags. Since the main intent of pattern discovery is to get a picture of the user's behavior, it does not make sense to include file requests that the user did not explicitly request.

CS Data Preparation

Data preparation includes the following activities:

- removing not relevant entries from log files;
- merging different events (URLs) that have the same result for user. For example, URL = .../admin/ means the same as URL = .../admins or URL = .../?cat=1 means the same as URL = .../about_company;
- listing all possible unique events (events with the same meaning are listed as one item) and all possible unique users. Replacing events with identificators;

- listing all user sessions (splitting all user events per user sessions. We defined new user session in the case there was more than 10 minutes interval between events of the same user). It is important to note that a user can perform several processes within one session;
- removing all user sessions that include only one activity, because these users have visited webpage accidentally and did not execute any process in the portal.

Before data preparation, there were ~2000 unique events and ~800 user sessions. After data preparation, there were 842 unique events. The result of data preparation is presented in Figure 1.

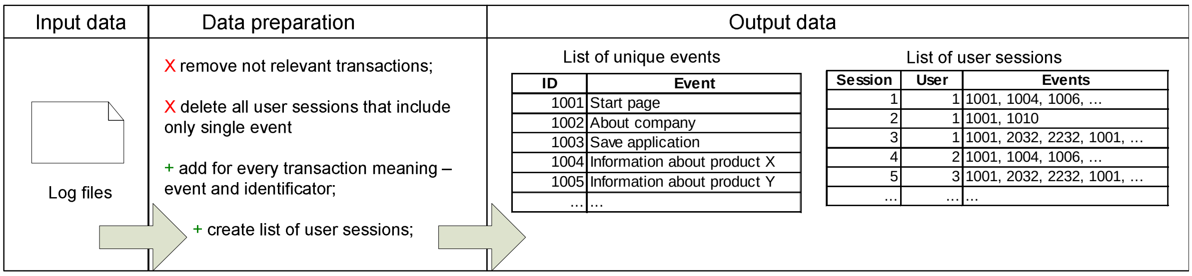

Figure 1. The data preparation process for CS

ES Data Preparation

Data preparation includes the following activities:

- listing all possible unique events (events with the same meaning are listed as one item) and all possible unique users. Replacing events with identificators.
- listing all user sessions (splitting all user events per user sessions). It is important to note that a user can perform several processes within one session.
- Removing all user sessions that includes only one event.

There are 29 unique events and 2325 user sessions. Result of data preparation is presented in Figure 2.

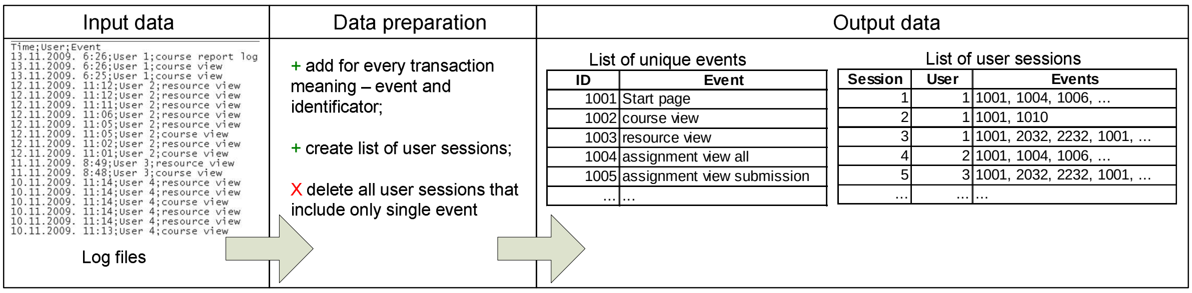

Figure 2. The data preparation process for ES

3.3 Finding and Filtering of Patterns

The pattern is defined as a sequence of events what is present within at least two sessions. For example, if one session contains events {A, B, C, D, E} and other session contains events {M, B, D, E, G} then the following sequence of events is present on both sessions {B, D, E}, and it is referred as to a pattern. These event items might not follow each other immediately and there could be events between them.

Patterns can be identified using different temporal data mining algorithms [8]. The Longest Common Sub-sequence algorithm is one of the most simple but also efficient algorithms. The longest common subsequence (LCS) problem is to find the longest subsequence L common to all sequences in a set of sequences (often just two). For the general case of an arbitrary number of input sequences, the problem is NP-hard. For the case of two sequences of T_i and T_j elements, the running time of the dynamic programming approach is $O\left(T_i \times T_j\right)$. L is determined using stepwise comparison of every two session or sequences of events. The following expression is evaluated on every step of the comparison:

$$L(S_1^i, S_2^j) = \begin{cases} 0 & if\ i = 0\ or\ j = 0 \\ L(S_1^i, S_2^j) \cup a_i & if\ a_i = b_j \\ L(S_1^i, S_2^{j-1}) & if\ a_i \neq b_j \wedge \left|L(S_1^i, S_2^{j-1})\right| > \left|L(S_1^{i-1}, S_2^j)\right| \\ L(S_1^{i-1}, S_2^j) & if\ a_i \neq b_j \wedge \left|L(S_1^i, S_2^{j-1})\right| < \left|L(S_1^{i-1}, S_2^j)\right| \end{cases} . \quad (3)$$

$L(S_1^i, S_2^j)$ denotes an intermediate longest common sequence, and S_1^i and S_2^j denotes *i*th and *j*th prefixes of sessions S_1 and S_2, respectively. If two intermediate longest common sequences are of the same length but not identical, then both are retained. LCS does not necessarily yield unique results, for example the longest common sequence of {*A*, *B*, *C*} and {*A*, *C*, *B*} is both {*A*, *B*} and {*A*, *C*}. Patterns are derived from all longest common sequences obtained during the stepwise comparison.

Patterns consisting of zero or one event are deleted (see figure 3). 0-item pattern means that there is nothing common between two sessions. 1-item pattern means that there is just one common event and that cannot be used as a basis for process oriented adaptation.

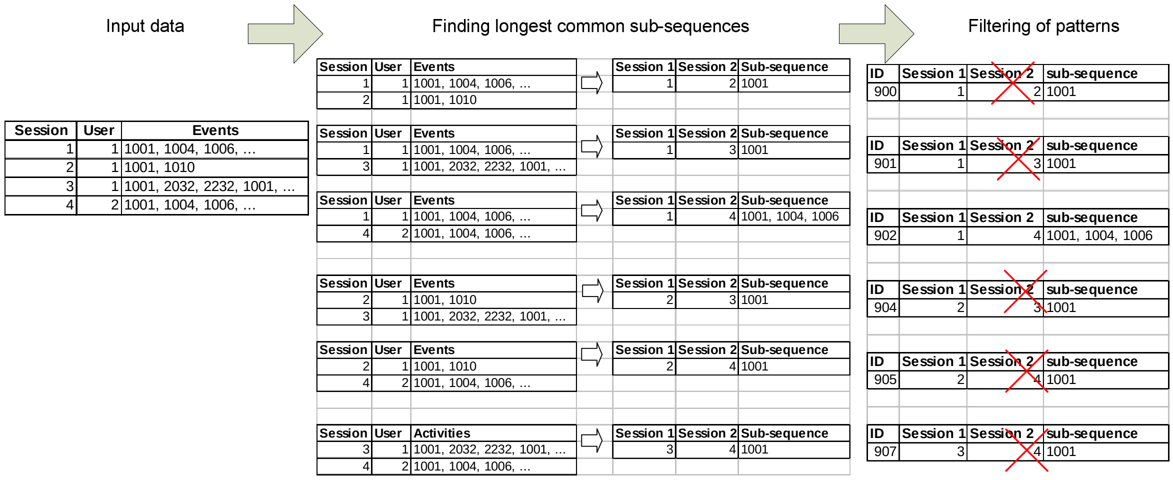

Figure 3. Filtering of patterns

3.4 Evaluation of Patterns

Patterns are evaluated using support and confidence measures as defined using Eq. 1 and Eq. 2, respectively. These measures are computed separately for general and individual patterns. Maximum possible support and confidence is 1 or 100%. If the support of pattern is closer to 1, then it is a popular pattern. If the confidence of pattern is closer to 1, then it is a precise pattern. The pattern is perceived as an efficient (i.e., it could be used efficiently for process adaptation) pattern if both support and confidence are high. A pattern efficiency chart is used to visualize pattern efficiency (Figure 4). The chart positions each discovered pattern according to its support and confidence levels. It is divided in four quadrants.

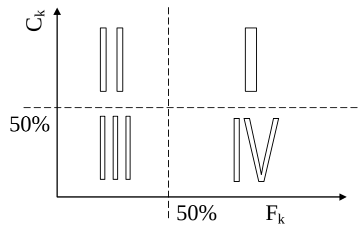

Figure 4. Efficiency of patterns

Each quadrant contains patterns with certain characteristics:

- Quadrant I – popular and precise patterns (confidence more than 50% and support more than 50%);
- Quadrant II – not so popular, but precise patterns (confidence more than 50% and support below 50%).
- Quadrant III – not so popular and not so precise patterns (confidence below 50% and support below 50%);
- Quadrant IV – popular, but not so precise patterns (confidence below 50% and support more than 50%);

It is obvious that the best patterns are within Quadrant 1 and not efficient patterns are within Quadrant 3. Therefore, patterns in each quadrant are counted separately for general and user patterns, and association between counts is used to check the overall hypothesis that personalized patterns are more efficient than general patterns.

4 Results

The methodology described above is used to analyze empirical information systems usage data obtained from CS and ES. Analysis results are presented for each information system separately.

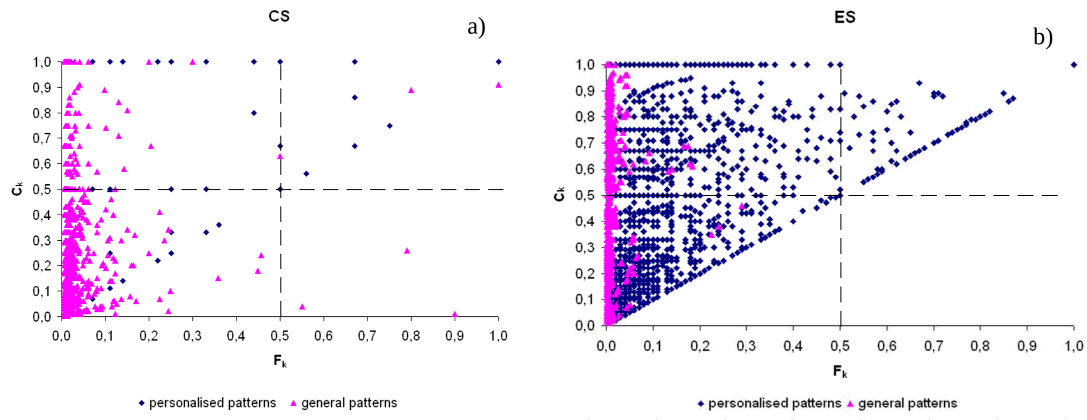

Figure 5. Pattern efficiency chart for (a) CS data and (b) ES data

Figure 5 shows the pattern efficiency charts for CS and ES data. One point in the charts can represent multiple patterns if their values of support and confidence numbers are the same. In the case of CS data, there is some contamination due to insufficient pre-processing of the log files. Visually, it can be observed that the general patterns are mainly located in the third quadrant while personalized patterns also frequent in the first and second quadrants.

In Figure 5a there is a point (personalized pattern) that has maximum support and confidence, in reality under this point are represented 75% of all personalized patterns. In this set are many patterns consisting of 2 events and some of these 2-item patterns are repeated by various users, but length of 'ideal' patterns are not limited only to 2 items. There are also patterns with length of 3, 4, 5 and 6 events. Some of these patterns are executed only by single user, others – by group of different users.

In Figure 5b someone could recognize that all personalized patterns are located on top of line where F_k=C_k. There is no special interpretation, just for every single personalized pattern Confidence always seems to be higher than support or equal to it. This tendency was not observed with general patterns.

Table 1. Cross-tabulation results

	CS data		ES data	
	General patterns	Personalized patterns	General patterns	Personalized patterns
Quadrant I	0.2	91.3	0.0	0.9
Quadrant II	22.0	7.4	65.5	79.2
Quadrant III	75.2	1.3	34.5	20.0
Quadrant IV	2.7	0.0	0	0.0
	Pearson Chi-Square = 11853, DF = 3, P-Value = 0.000		Pearson Chi-Square = 667, DF = 3, P-Value = 0.000	

Numerical evaluation of pattern efficiency is obtained by cross-tabulation (Table 1) showing percentage of patterns falling in each quadrant for general patterns and for personalized patterns. It can be observed that 0,2% of general patterns and 91,3% of individual patterns are precise and popular at the same time for CS data (Quadrant I). At the same time 75,2% of general patterns and only 1,3% of individual patterns fall in Quadrant III (least desirable situation). It means that data from these log files confirm our assumption that individual patterns are more effective. Additionally, personalized patterns have very high efficiency and almost all patterns are very precise and popular. That means that if users return to webpage, their activities are the same almost in all sessions. For the ES data, it can be observed that 0% of general patterns and 0.9% of individual patterns are precise and popular at the same time. The majority of both general and personalized patterns belong to Quadrant II while fewer personalized patterns belong to Quadrant III. That means that also personalized patterns in discovered from the ES data are slightly more efficient than general patterns. The Chi-square test of cross-tabulation data confirms that the position of patterns in the pattern efficiency chart depends upon pattern personalization. Given that majority of personalized patterns fall in Quadrant I, personalized patterns are more efficient than general patterns at least in the case of CS data.

Table 2. Descriptive statistics of the support and confidence measures

	F		*C*	
	General patterns	Personal patterns	General patterns	Personal patterns
CS data				
Mean	0.03	0.85	0.39	0.96
St. Dev.	0.06	0.27	0.34	0.14
Median	0.01	1.00	0.30	1.00
ES data				
Mean	0.01	0.06	0.65	0.75
St. Dev.	0.02	0.08	0.29	0.31
Median	0.00	0.04	0.67	1.00

Non-parametric Mann-Whitney test was used to compare support and confidence values for personalized and general patterns to test H1 and H2 presented in Section 2. Probability that the median value for personalized patterns is larger than median value for generalized patterns approaches zero for both measures. Therefore, both hypotheses are accepted implying that the personalized patterns are observed more frequently and their predictive power is stronger than that of general patterns. It has to be noted more frequent observation of personalized patterns could have been expected because of size differences between $\mathbf{S}$ and $\mathbf{S}_u$. However, the magnitude of differences suggests that using general patterns as a basis of adaptation could be misleading. The finding that personalized patterns have higher confidence is not so obvious and suggests that individual users have strong tendency to use information systems in a predictable manner what could be exploited in design of adaptive information systems.

There is a large difference of pattern support level between CS and ES data. That could be explained by using process groups in CS data while process groups have not been used in ES data. It appears that users use only a limited number of processes. Therefore, patterns belonging to one process group are not likely to occur in sessions during which user primarily considers processes from unrelated process groups.

5 Related work

Analysis of transaction logs had been widely used to understand user behavior within the system, e.g. to characterize user search behavior in information retrieval systems [7, 9, 15], to do process mining to improve quality of services and processes [12, 13] or for extracting business rules from information systems [14]. Most of the studies concentrate on mining usage of web systems. The overall process of web usage mining is generally divided into two main tasks: data preprocessing and pattern discovery. Data mining methods employed during pattern discovery include association rule mining (e.g. [2]), sequential pattern discovery (e.g. [6]), clustering (e.g. [5]) and classification (e.g. [1, 10]) or mix of the aforementioned methods (e.g. [18]). Usage patterns extracted from web data can be applied to a wide range of applications such as web personalization, system improvement, site modification, business intelligence and usage characterization [16].

According to Gerry [4] and Wang [19], the accuracy and coverage rate of association mining techniques is usually quite low. They applied the association mining over all users‘ navigation sessions to establish a knowledge model that predicts users‘ next request. Their results also show that the sequence mining method produces higher accuracy than association mining. The association patterns do not perform well in prediction of future navigation patterns due to the low matching rate of prediction rules. Some improvements are made, when instead of performing the association mining task over all users‘ navigation sessions, users are first clustered so that users in each cluster demonstrate shared navigation characteristics [18].

However, none of the approaches analyzes personal dimension of discovered patterns. The navigational behavior of the current user is matched with the ones of other users with similar profiles, so the adaptation is done according to the profiles of other similar users, not according to patterns of a particular user.

6 Conclusion

The analysis of empirical information systems usage data has been performed to highlight differences between general system usage and behavior of individual users who might prefer their personalized ways information system usage. The empirical data obtained from log files of CS and ES confirm the hypotheses that personalized pattern are observed more frequently than general patterns and that personalized patterns have higher confidence or predictive power. From the perspective of the proposed development of process-oriented adaptive information systems, these findings imply that users are more likely to accept adaptive systems that are based on personalized patterns and that the usage of personalized patterns could result in more accurate adaptive behavior of the system (e.g., system’s recommendations about further steps of the process execution would be more accurate). Acceptance of both hypothesis and cross-tabulation results show that overall efficiency of personalized patterns is better than that of general patterns.

However, it can be observed that efficiency of patterns varies significantly even between just two empirical data sets. That implies that expected gains from adaptivity also would vary significantly. Low level of pattern support in the case of ES implies that an adaptive system would have to support very large number of different process variations what could be technically challenging. It has been identified that process groups might have significant role in prediction of pattern efficiency, it is stated also in [10], where term '*content*' refer to the same meaning as '*process group'* in this paper.

References

[1] **Cho, Y.H., Kim, J.K.,** Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce, *Expert Systems with Applications*, vol.26, pp.233-246, 2004.

[2] **Cooley, R., Mobasher, B., Srivastava, J.,** Data preparation for mining world wide web browsing patterns, *Knowledge and Information Systems* 1 (1), 1999.

[3] **Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C**., Introduction to Algorithms (2nd ed.), *MIT Press and McGraw-Hill*, pp.350-355, 2001.

[4] **Gery, M., Haddad, H.,** Evaluation of Web usage mining approaches for user's next request prediction, *Proceedings of the fifth ACM international workshop on Web information and data management*, pp.74-81, 2003.

[5] **Girardi, R., Marinho, L.B., De Oliveira, I.R**., A system of agent-based software patterns for user modeling based on usage mining, *Interacting with Computers*, Volume17, pp.567-591, 2005.

[6] **Huang, Y.M., Kuo, Y.H., Chen, J.N., Jeng, Y.L**., NP-miner: A real-time recommendation algorithm by using web usage mining, *Knowledge-Based Systems*, vol.19, pp.272-286, 2006.

[7] **Jones, S., Cunningham, S.J., McNab, R. and Boddie, S**., A transaction log analysis of a digital library, *International Journal on Digital Libraries*, Volume 3, Number 2/August, 2000.

[8] **Laxman, S**., **Sastry, P.S.,** A survey of temporal data mining, *Sadhana*, Volume 31, Number 2/April, 2006.

[9] **Lazarinis, F., Vilares, J. and Efthimiadis, E.N.,** Current research issues and trends in non-English Web searching, *Information Retrieval*, volume 12, Number 3/June, 2009.

[10] **Liu, H., Keselj, V.,** Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users' future requests, *Data and Knowledge Engineering*, vol.61, pp.304-330, 2007.

[11] **Maglio, P.P., Campbell, C.S., Barrett, R., Selker, T.,** An architecture for developing attentive information systems, Knowledge-Based Systems, Volume 14, pp.103-110, 2001.

[12] **Măruşter, L., Beest, N.R.T.P.,** Redesigning business processes: a methodology based on simulation and process mining techniques, *Knowledge and Information Systems*, Volume 21, Number 3/December, 2009.

[13] **Rozinat, A., Mans, R.S., Song, M. and Aalst, W.M.P**, Discovering colored Petri nets from event logs, *International Journal on Software Tools for Technology Transfer (STTT)*, Volume 10, Number 1/January, 2008.

[14] **Shao, J., Pound, C.J**., Extracting Business Rules from Information Systems, *BT Technology Journal*, Volume 17, Number 4/October 1999.

[15] **Smeaton, A.F., Callan, J.,** Personalisation and resommender systems in digital libraries, *International Journal on Digital Libraries*, Volume 5, Number 4/August, 2005.

[16] **Srivastava, J., Cooley**, R., Deshpande, M., Tan, P., Web usage mining: discovery and applications of usage patterns from Web data, *SIGKDD Explorations*, 1 (2), 2000.

[17] **Šūpulniece, I., Grabis, J**. User Modelling Driven Adaptive Enterprise Applications, *5th International Conference on Enterprise Systems, Accounting and Logistics (5h ICESAL '08)* 7-8 July 2008, Crete Island, Greece.

[18] **Wang, F.H., Shao, H.M.,** Effective personalized recommendation based on time-framed navigation clustering and association mining, *Expert Systems with Applications*, 27, pp.365-377, 2004.

[19] **Wang, F.H., Thao, S.M.,** A study on personalized Web browsing recommendation based on data mining and collaborative filtering technology, *Proceedings of national computer symposium*, Taiwan, pp.18-25, 2003.