Intelligent Information Personalization Leveraging Constraint Satisfaction and Association Rule Methods

Syed Sibte Raza Abidi and Yan Zeng

Faculty of Computer Science, Dalhousie University,
Halifax, B3H 1W5, Canada
{sraza, yzeng@cs.dal.ca}

Abstract. Recommender systems, using information personalization methods, provide information that is relevant to a user-model. Current information personalization methods do not take into account whether multiple documents when recommended together present a factually consistent outlook. In the realm of content-based filtering, in this paper, we investigate establishing the factual consistency between the set of documents deemed relevant to a user. We approach information personalization as a constraint satisfaction problem, where we attempt to satisfy two constraints—i.e. user-model constraints to determine the relevance of a document to a user and consistency constraints to establish factual consistency of the overall personalized information. Our information personalization framework involves: (a) an automatic constraint acquisition method, based on association rule mining, to derive consistency constraints from a corpus of documents; and (b) a hybrid of constraint satisfaction and optimization methods to derive an optimal solution comprising both relevant and factually consistent documents. We apply our information personalization framework to filter news items using the Reuters-21578 dataset.

1 Introduction

Information seekers are different in nature in that they manifest different information seeking behavior, therefore their information seeking experience and outcome should not only be unique but it should be tailored to their individual persona, purpose, interests, educational backgrounds, demographics and preferences. Information Personalization (IP) research purports strategies to either filter or adapt information items based on both the user's characteristics and information retrieval criterion [1, 2, 3]. The ensuing information personalization systems employ adaptive hypermedia, information retrieval and artificial intelligence methods to (a) formulate a user-model and (b) leverage this user-model to personalize the information to be recommended to an individual user. From an AI perspective, a variety of techniques have been employed for pursuing IP. Foltz used latent semantic indexing (LSI) to perform information personalization [4]; Mooney and Roy developed a book recommending system based on a Bayesian text classifier [5]; Malone et al built a rule-based system to filter e-mail messages [6]; Jennings and Higuchi helped users get better access to news service using neural-networks [7]; Desjardins and Godin use genetic algorithms

L. Lamontagne and M. Marchand (Eds.): Canadian AI 2006, LNAI 4013, pp. 134 – 145, 2006.

for personalization [8]. However, pursuing IP as a constraint satisfaction problem is a novel approach.

Notwithstanding the efficacy of intelligent IP systems to determine the relevance of the recommended information item towards a user-model, it can nevertheless be argued that the underlying information personalization mechanism do not account for the factual consistency between the recommended information items—i.e. whether multiple recommended information items when presented together present a consistent outlook or inadvertently lead to a contradictory outlook. We believe that whilst two information items may be relevant to the user model, there may be instances when their simultaneous presentation to the user can potentially lead to a situation whereby one information item is stating a certain fact/recommendation whilst the other information item maybe contradicting the same fact/recommendation. Alternatively, users may seek information items that present divergent views in which case factually inconsistent information items need to be presented to the user. In each case, the requirement is to establish the factual similarity/dissimilarity between two information items.

We approach IP as a constraint satisfaction problem. Intuitively speaking, the problem of information personalization entails the satisfaction of two different constraints for each information item: (a) *relevancy constraints* to establish the relevance of the document to the user; and (b) *consistency constraints* to establish the factual consistency between the selected documents. Our approach to IP involves the satisfaction of the abovementioned constraints such that: (i) given a large set of documents we select only those documents that correspond to the user-model; (ii) given the selected user-compatible documents, we retain only those documents that cumulatively present a level of factually consistency as specified by the user; and (iii) we attempt to maximize the information coverage of the personalized information by selecting the largest possible set of documents that satisfy the above two constraints. In our work IP is achieved without deep content analysis, rather by leveraging the pre-defined classification of documents in terms of topics.

In this paper, we build on our previous work on IP [9, 10] by extending it in terms of (a) an automatic constraint acquisition method based on association rule mining [11] to derive consistency constraints from a corpus of documents. This current method eliminates the need for acquiring consistency constraints from domain experts which was previously viewed as a bottleneck; (b) adding more flexibility to the constraint satisfaction framework by solving IP as an Over-constrained CSP through a hybrid of partial constraint satisfaction and optimization methods; and (c) a user preference setting mechanism whereby users can set the personalization criteria, such as tolerance to inconsistency or degree of information comprehensiveness in line with their information needs. We demonstrate the working of our IP framework for news item selection for a personalized news delivery service using the Reuters-21578, Distribution 1.0 data-set.

2 Specification of an IP Problem

Computationally, constraint satisfaction methods allow the efficient navigation of large search spaces to find an optimal solution that entails the assignment of a value from its domain to every problem variable, in such a way that every constraint is satisfied. This may involve finding (a) just one solution with no preferences, (b) all solutions, or (c) an optimal solution given some objective function [12, 13, 14]. In our work, the problem of IP is specified as follows:

2.1 User-Model

The user-model characterizes the user in terms of: (a) user's *interests* represented as a list of topics, (b) user's *tolerance* towards inter-document inconsistency, and (c) user's *preference* towards the coverage of the solution—i.e. whether the solution should satisfy all user-interests or instead it should satisfy all consistency constraints.

2.2 Information Items

The information items (i.e. documents) comprise two sections: (a) *Content* section that contains the actual information; and (b) *Context* section that contains a list of topics categorizing the document. During the IP process, the topics in the context section are compared with the topics mentioned within user-model to determine the relevance of a document to a particular user.

2.3 Information Personalization Constraints

Two types of constraints are used to pursue IP: (a) *Relevancy constraints* to ensure that the selected documents are relevant to the user's interest as specified in the user-model; and (b) *Consistency constraints* to (i) ensure that the personalized information is factually consistent. This is achieved through *negative consistency constraints,* which define what pairs of topics cannot co-exist together. Negative consistency constraints are represented as the tuple *nc* (*topic1, topic2, degree*), where *degree* is the degree of inconsistency between the two topics. Two documents cannot be simultaneously presented to the user if the topics they represent cannot coexist; and (ii) to maximize the coverage of the personalized information. This is achieved through *positive consistency constraints*, which define what topics' if simultaneously presented would likely be of interest to the user. Positive consistency constraints are represented as the tuple *pc* (*topic1, topic2, degree*), where *degree* is the degree of similarity between the two topics. For example, recently in the news the topics *Ice-Skating* and *Winter Olympics 2006* appeared quite frequently, thus suggesting a positive consistency constraint between *Winter Olympics 2006* and *Ice-Skating.* Such a constraint can be used to recommend additional information about *Winter Olympics 2006* if the user is interested in *Ice-Skating* and vice versa.

2.4 Information Personalization Requirements

Given a user-model, a corpus of documents and a set of constraints, our solution to an IP problem needs to address the following requirements:

1. The personalized information should be relevant to the interests of the user. The user may choose the degree of relevance to include either all or a partial list of topics of interest.

2. The personalized information should be factually consistent—i.e. the set of documents being presented to the user should mutually satisfy the consistency constraint.
3. The IP process should attempt to find the largest set of consistent documents in terms of the coverage of topics defined in the user-model.

2.5 Defining IP as a Constraint Satisfaction Problem

In our constraint satisfaction approach for information personalization, the topics representing the user's interest are viewed as variables, and domains of the variables comprise any combination of available documents. Requirement 1 can be solved as a unary constraint to the variables and represented by constraint c_1. Requirement 2 can be represented by a unary constraint c_2 and a binary constraint c_3. Requirement 3 can be addressed through an objective function O. c_1, c_2, c_3 and O are explained below.

We define our IP problem as $P\ (V, D, C, O)$.

- Variable set $V = \{v_1, v_2, \ldots, v_n\}$, where n is the number of topics of a user's interest; v_i, $1 \le i \le n$, represents the i^{th} topic of a user's interest.
- Domain set $D = \{d_1, d_2, \ldots, d_n\}$; d_i, $1 \le i \le n$, represents the domain of v_i. Suppose $s = \{t_1, t_2, \ldots, t_m\}$ is a set consisting of all documents, then d_i is the power set of s without the empty set ø. E.g. If $\{t_1, t_2\}$ is the set of documents, the domain of the variable will be $\{\{t_1\}, \{t_2\}, \{t_1, t_2\}\}$.
- Constraint set $C = \{c_1, c_2, c_3\}$; $c_1 = rel(v_i)$, where $1 \le i \le n$, is a unary constraint, and means the value of v_i must be relevant to users' interest (Requirement 1). Suppose v_i represents the i^{th} topic of a user's interest, and the domain of v_i is $\{\{t_1\}, \{t_2\}, \{t_1, t_2\}\}$. By checking the topics of t_1 and t_2, we know t_1 is relevant to the i^{th} topic of the user's interest, but t_2 is not. To satisfy c_1, $\{t_2\}$ and $\{t_1, t_2\}$ will be removed from the domain of v_i. $c_2 = con1(v_i)$, where $1 \le i \le n$, is a unary constraint, and means the documents assigned to v_i must be consistent to each other (Requirement 2). Suppose the system is trying to assign $\{t_1, t_2\}$ to v_1. To decide whether c_2 is satisfied or not, we can check the consistency between t_1 and t_2. Suppose t_1 presents topics 'acquisition' and 'stocks', and t_2 presents topics 'acquisition' and 'gold'. We take one topic from t_1 and t_2 respectively to form pairs of topics ordered alphabetically. Then we get four pairs - *(acquisition, acquisition)*, *(acquisition, gold)*, *(acquisition, stocks)* and *(gold, stocks)*. We check these four pairs against the effective negative consistency constraints, and find that *(acquisition, gold)* triggers a negative constraint. So we know c_2 is violated and the assignment fails. $c_3 = con2(v_k, v_j)$, where $k \neq j$ and $1 \le k, j \le n$, is a binary constraint, and means the value of v_k and v_j must be consistent to each other (Requirement 2). When checking c_3, we take a document from the value of both variables to form pairs of documents. If any pair is inconsistent, c_3 is violated.
- $O = \Sigma_i\ (n_i * weight_i)$ is the objective function, where i is a member of the set of satisfied positive consistency constraints--S. n_i is the time the constraint i is satisfied. $weight_i$ is the correlation value of the constraint i. The target is to find a complete valuation that maximizes the objective function. This function will be used in step3 (coverage maximization) of our CSP process solving.

- From the above specification, it can be seen that our IP problem is an Over-constrained CSP (OCSP)—i.e. a complete valuation that satisfies all hard constraints cannot be guaranteed—because the settings of the user's information personalization preferences may lead to the non-satisfaction of the negative consistency constraints. In this case, (a) if a user prefers *maximum coverage of the topics of interest* then the solution that covers the largest possible number of topics of interest whilst violating the least number of negative consistency constraints will be selected, and (b) If the user prefers a certain degree of *consistency in the adapted information* then the solution will allow only the corresponding violation of negative consistency constraints. In order to address OCSP, we have modified our CSP as follows: (i) add the empty set ϕ to the domain of variables; and (ii) add a collection of constraints, $c_4 = \{no\_empty(v_i), 1 \le i \le n\}$. It means the empty set is a variable's last choice. Now the constraint set $C = \{c_1, c_2, c_3, c_4\}$.

3 Constraint Satisfaction Based IP Framework

Our IP framework performs two related functions: (a) given a corpus of documents it automatically finds the consistency constraints; and (b) given a user-profile it generates an information personalization solution. The functional steps (in shaded boxes) and the technical methods used in our IP framework are illustrated in Figure 1.

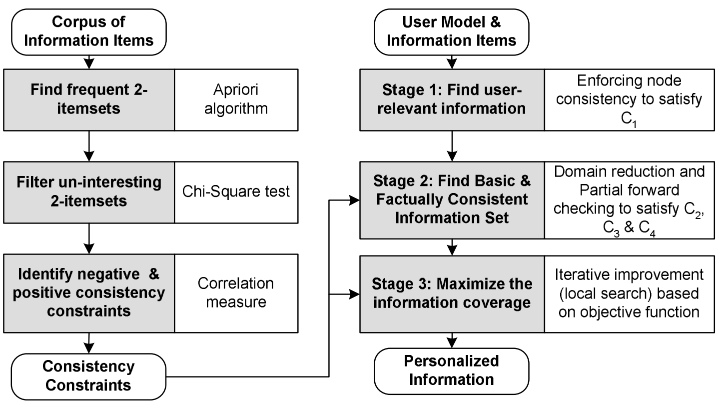

Fig. 1. The functional steps and the corresponding methods used in our IP framework

Our information personalization strategy works in three main stages: (1) Find all the information items relevant to the user-model; (2) Find a simplified solution whereby each user interest is accounted for by a single information item, whilst ensuring factual consistency between the selected information items; (3) Use the simplified solution as the basis to maximize the scope of the solution by including more information items that satisfy both relevance and consistency constraints.

3.1 Our Approach for Consistency Constraint Acquisition

One problem that we faced in our previous work was the acquisition of consistency constraints from domain experts. The literature is inconclusive in this regard. Padmanabhuni et al [15] suggest a framework for learning only positive constraints for discrete domain; O'Sullivan et al [16] use an interactive approach to acquire constraints from users by searching through a 'hypothesis space' of constraints.

In our current work, we addressed this problem by acquiring consistency constraints directly from the given corpus of information items (with pre-assigned topics) by using the association rule-mining approach [5]. The premise of the approach is that when information is composed it entails some inherent relationships between discussion topics that can meaningfully co-occur within a given document. Such relationships between topics are largely determined by the authors' working knowledge. We leverage these intrinsic relationships between topics to establish consistency constraints such that the frequency of co-occurrence of information topics may reflect the degree of consistency between the topics. We treat topics as items in the *Apriori* rule association method to find 2-itemsets [5]. We select the 2-itemsets with high support value and calculate the correlation between the two items as $corr(A,B) = \dfrac{p(AB)}{p(A)p(B)}$. The correlation value is used to distinguish between positive and negative consistency constraints as follows:

- If $0 < corr(A, B) < 1$, A and B are correlated negatively it means these two topics are inconsistent to each other, so a *negative consistency constraint* can be established between these two topics.
- If $corr(A, B) > 1$, A and B are positively correlated, and they encourage the co-occurrence of each other, so a *positive consistency constraint* is found between these two topics.
- If $corr(A, B) = 1$, A and B are independent to each other.

After our experiments with the Reuters-21578 dataset, we acquired 913 frequent 2-itemsets We used the Chi-Square statistical significance test to measure the interestingness of the 2-itemsets [17], where the Chi-Square significance level was set

Table 1. Illustration of some 2-itemsets and their selection as consistency constraints

Topic1	Topic2	Frequency	Correlation	Chi_Square	Action
crude	natural-gas	81	11.171	803.615	**Positive Constraint**
rice	wheat	20	11.089	189.774	**Positive Constraint**
livestock	soy-meal	3	11.079	.345	Removed
…	…	…	…	…	…
grain	trade	20	.657	3.984	**Negative Constraint**
…	…	…	…	…	…
coffee	crude	3	.371	3.433	Removed
acquisition	natural-gas	10	.357	14.914	**Negative Constraint**
…	…	…	…	…	…
acquisition	money-fx	1	5.797E-03	233.783	**Negative Constraint**

to 95% and we acquired a smaller-sized but high quality set of consistency constraints. The consistency constraints were sub-divided into positive and negative consistency constrains based on their correlation values (as shown in Tables 1 and 2).

Table 2. Final distribution of the consistency constraints

Positive/Negative	2-item rules	Interesting 2-item rules
Positively correlated	768	120
Negatively correlated	145	57
TOTAL	913	177

3.2 Solving the Constraint Satisfaction Problem for Information Personalization

We highlighted earlier that our information personalization is an OCSP, and hence its solution can be viewed as a partial constraint satisfaction problem (PCSP) in which a complete valuation is made with some constraints unsatisfied, and the valuation with the smallest distance is selected as the final solution. The distance can be defined as the number of constraints violated by a valuation [18]. Our strategy to solve the PCSP is explained using an exemplar user profile (in Table 3) and dataset (in Table 4).

Table 3. User profile used in the working example

Component	Value
Interests	Acquisition, Gas, Income, Jobs
Tolerance	20% factual inconsistency
Preference	Satisfy all consistency constraints

Table 4. Dataset used for the example

News Item	Topics	News Item	Topics
t_1	acquisition	t_9	jobs
t_2	acquisition, crude, nat-gas	t_{10}	bop, cpi, gnp, jobs
t_3	acquisition, gold, lead, silver, zinc	t_{11}	jobs, trade
t_4	gas	t_{12}	gnp, jobs
t_5	CPI, crude, fuel, gas, nat-gas	t_{13}	acquisition
t_6	fuel, gas	t_{14}	fuel, gas
t_7	crude, gas	t_{15}	jobs, trade
t_8	GNP, income, ipi, retail, trade		

From the user's interests we get four variables, each representing a topic of the user's interest. We refer to these variables as v_{acq}, v_{gas}, v_{income} and v_{jobs}. The domain of these four variables is the power set of the 15 news items that are shown in Table 4.

Step 1: Filter User-Relevant Information

The first step involves finding all the documents that correspond to the user's interest as per requirement 1of the information personalization specification. This involves the satisfaction of the relevancy constraint by enforcing node consistency to satisfy the unary constraint $c_1 = rel(v_i)$ by comparing the topics of the various documents against a user's interest as noted in the user-model. The node representing the variable v in a constraint graph is node consistent if for every value x in the current domain of v, each unary constraint on v is satisfied. Functionally, if the variable v_{acq} has a value (i.e. news item) that is not equal to the topic '*acquisition*' (one of the four interests of the user) then the value will be filtered out from v_{acq}'s domain. The same process is repeated for v_{gas}, v_{income} and v_{jobs} in case of our working example. At the end of step one we get user-relevant news items for each variable as shown in the second column of Table 5. For example, the relevant set of v_{acq} is found to be $\{t_1, t_2, t_3, t_{13}\}$ and for v_{gas} the relevant set is $\{t_4, t_5, t_6, t_7, t_{14}\}$. After step1, the domain of a variable is the power set of its relevant set, i.e. the domain of v_{acq} is the power set of $\{t_1, t_2, t_3, t_{13}\}$.

Table 5. User relevant items for the variables

Variable	Retained Relevant item	Removed Relevant item	Variable	Retained Relevant item	Removed Relevant item
v_{acq}	t_1, t_2, t_3	t_{13}	v_{income}	t_8	
v_{gas}	t_4, t_5, t_6, t_7	t_{14}	v_{jobs}	$t_9, t_{10}, t_{11}, t_{12}$	t_{15}

Step 2: Find the Basic Information Set

At the end of stage 1, the size of the set of user-relevant items is typically quite large, and likewise the resulting power set is quite large. We feel that in such a situation it is unwise to use systematic methods to solve OCSP because we will probably just be able to find a partial solution for the problem, whereas there may exist the possibility to completely solve the problem—i.e. the adapted information is imperfect and does not meet a user's requirements as perfectly as it is possible. In order to personalize the information with respect to all the constraints in the constraint set C, we attempt to solve a simplified version of the original problem in order to find out: (a) whether the problem can be satisfied completely or not? If not, what is the least number of violated constraints? and (b) what feasible solutions can be used as the starting point for the optimization process in order to maximize the coverage of the personalized information. To answer the above questions, we pursue domain reduction—i.e. eliminate some elements from the domain of variables to make it feasible to search the solution space systematically to find a *basic information set*. A solution is called *basic information set* if (i) each user interest is assigned at most one information item; and (ii) it violates the least number of consistency constraints; and (iii) least number of user-interests have no information item. Domain reduction is done in three steps.

First, we delete duplicate items from the set of user-relevant documents. If a document represents a group of topics that are also represented exactly by other documents, then we keep one document and remove the others.

Second, we delete values with multiple elements from the domain of variables because values with multiple elements can unnecessarily violate more consistency constraints. This enables the domain size of a variable with k relevant items to be reduced from 2^k to k+1.

Third, we delete dominating values (sets) from the domain. If the topic set of item t_1 is a subset of the topic set of item t_2, we say t_2 dominates t_1, and t_2 is a dominating item. If a value contains only dominating items, it is a dominating value. Since a dominating item comprises extra topics it offers a stronger likelihood to violate more consistency constraints as compared to the item that it dominates. It may be noted that if t_2 dominates t_1 and t_1 is inconsistent with t_3, t_2 is inconsistent to t_3 too. Hence, if we have checked the consistency between the value of $\{t_1\}$ and $\{t_3\}$, we do not need to check the consistency between $\{t_2\}$ and $\{t_3\}$ any more. So we can eliminate all dominating values from the domain without changing the characteristics of the problem. Here, we just show for v_{acq} the details of deleting multi-elements and dominating values in Table 6, and the resulting domain of the four variables is shown in Table 7.

Table 6. Deleting multi-element and dominating values

Retained	Removed (dominating)	Removed (multi-elements)
Ø, $\{t_1\}$	$\{t_2\}$,$\{t_3\}$	$\{t_1, t_2\}$,$\{t_1, t_3\}$,$\{t_2, t_3\}$,$\{t_1, t_2, t_3\}$

Table 7. The domain of the variables

Variable	Domain	Variable	Domain
v_{acq}	{ ø , $\{t_1\}$}	v_{income}	{ ø , { t_8}}
v_{gas}	{ ø , { t_4}}	v_{jobs}	{ ø , { t_9}}

Step 3: Establish Factual Consistency of User-Relevant Information

This step involves establishing the factual consistency between the selected information. After domain reduction we have managed to simplify the solution space to apply a variant of *branch and bound method*—i.e. Partial Forward Checking (PFC)—that systematically searches for the solutions by satisfying the constraints c_2, c_3 and c_4. PFC being a variant of forward checking has been shown to perform better than most systematic methods used to solve PCSP [12]. It may be noted that in comparison with step1, which can be realized quite efficiently, and step4, which can be terminated at any time according to the availability of resources, step3 involves a systematic search and hence is the key to the success of the whole process of information personalization. To ensure this we conducted compared variants of PFC.

We apply PFC algorithm to our PCSP using two different distances: (i) the number of variables assigned to the empty set (violating c_4), referred to as d_{empty} and (ii) the number of times the negative consistency constraints are violated, referred as $d_{violation}$. If a user's preference is 'Satisfy all topics of interest', $d_{violation}$ will be used; otherwise, d_{empty} will be used. For our example, the results of using PFC is shown in Table 8.

Table 8. Factually consistent solutions after domain reduction

Solution	acquisition	gas	income	jobs
1	$\{t_1\}$	$\{t_4\}$	Ø	$\{t_9\}$
2	Ø	$\{t_4\}$	$\{t_8\}$	$\{t_9\}$

It may be noted that $\{t_1\}$ and $\{t_8\}$ are inconsistent to each other because of the effectiveness of *nc*(acquisition, trade, 0.034). For both solutions, the distance is

$d_{\text{empty}} = 1$. It means there is one topic of interest left empty. And this calculated distance is the minimum distance that solutions to the original problem can achieve.

Step 4: Maximize Information Coverage

In this step we attempt to maximize the information coverage of the solution obtained in step 3. Note that the solution at this stage contains at most one information item for every topic defined in the user's interest. This condition is in line with requirement 3 of our information personalization specification. We use local search based optimization techniques to improve the solution by assigning values with more elements (information items) to variables (topics of a user's interest) whilst maintaining the factual consistency.

The iterative improvement method [19] used works as follows: First, it sets the solution at step 3 as the current solution and then searches the current solution's neighborhood for a better solution. If there is such a solution, the current solution is set to this 'improved' solution, and the search goes on. Else, the current solution is returned as the result of optimization. The neighborhood of the current solution consists of all solutions whose difference from the current solution is just the value of one variable. Two criteria are used to determine which solution is better: (1) higher value of the objective function, i.e. a higher sum of degrees of the satisfied positive consistency constraints. The positive consistency constraints are checked in the same way the negative consistency constraints are checked, i.e. first construct item pairs from assigned values, then construct topic pairs from item pairs, and finally check topic pairs against positive consistency constraints. The optimization round using this criterion is called *positive consistency round*; and (2) higher number of information items. The optimization round using this criterion is called *cardinality round*.

For optimization purposes, the non-null variables in the factually consistent solution (i.e. Table 8) are restored with their original domain representing information items corresponding to the user's interests (as shown in Table 5). For instance, the domain of v_{acq} is restored to be the power set of $\{t_1, t_2, t_3\}$.

The optimization results (shown in Table 9) lead to two solution—i.e. solution3 and solution4. However, solution4 has the higher objective function value and hence is designated as the final optimized solution. Finally, the information items comprising solution4 will be presented to the user as the information personalization solution based on his/her user-model.

Table 9. Final optimized solutions

Solution	acquisition	gas	income	jobs	Objective function
3	$\{ t_1\}$	$\{ t_4, t_6\}$	NULL	$\{ t_9, t_{10}, t_{12}\}$	45.28
4	**NULL**	$\boldsymbol{\{ t_4, t_6\}}$	$\boldsymbol{\{ t_8\}}$	$\boldsymbol{\{ t_9, t_{10}, t_{11}, t_{12}\}}$	**121.65**

4 Evaluations of Variants of Partial Forward Checking

In general, variable and value ordering heuristics are effective in improving efficiency of systematic search methods. In this section we compare the performance of partial forward checking (basic_pfc), partial forward checking with variable ordering (order_pfc), partial forward checking with variable and value ordering (full_pfc) in terms of the number of constraint checks that is a standard measure of efforts for CSP algorithms. The basic branch and bound method was tested to compare it against PFC. The variable ordering heuristic used in this evaluation is the smallest-domain heuristic [20]. The value ordering heuristic used in this evaluation is to select first the value with minimal inconsistency count [20].

For Reuters-21578 dataset and a list of topics of a user's interest, we compare the performance of these algorithms by varying the user's preference (Fig. 2) and tolerance (Fig. 3). From our experiments we note that any variant of PFC performs better than branch and bound method. Furthermore, the full_pfc always gave the best performance which vindicates are decision to use PFC in step 3 for establishing factual consistency.

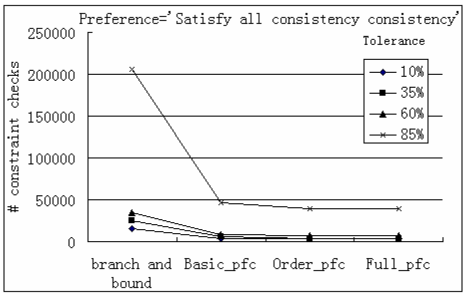

Fig. 2. Performance of algorithms (satisfying all consistency constraints)

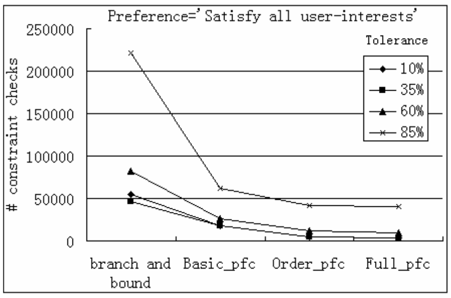

Fig. 3. Performance of algorithms (satisfying all user-interests)

5 Concluding Remarks and Future Work

Viewing information personalization as a constraint satisfaction problem offers an interesting AI based perspective to an information retrieval issue. We have demonstrated the successful application of a hybrid of constraint satisfaction methods that offer personalized information that is based on user's interests and personalizetion preferences. Our information personalization strategy makes it possible to find better sub-optimal solutions by combining systematic search and local search for the information personalization problem, and we believe this approach can be applied to other fields as well. In this work, we additionally addressed the core issue of constraint acquisition from the domain knowledge as opposed from domain experts. Our current association rule based approach works with the a priori defined classification of the documents. In future we plan to analyze the content of the document, as opposed to meta-level topics, to establish richer consistency constraints using automated text categorization techniques involving learning mechanisms.

References

1. Belkin N.J., Croft W.B. Information personalization and information retrieval: Two sides of the same coin? *Communications of the ACM*, 35(12), 1992, 29-38.
2. Baeza-Yates, R., Ribeiro-Neto, B. *Modern Information Retrieval*, Addison-Wesley, 1999.
3. Hanani, U., Shapira, B., Shoval, P. Information Filtering: Overview of Issues, Research and Systems. *User Modeling and User-Adapted Interaction,* 11, 2001, 203-259.
4. Foltz, P.W. Using latent semantic indexing for information filtering. In *ACM SIG-OIS*, 1990, 40-47.
5. Mooney, R.J., Roy, L. Content-based book recommending using learning for text categorization. In *Proceedings of the 5th ACM Conference on Digital Libraries*, June 2000, San Antonio, Texas, USA. 195-204.
6. Malone, T.W., Grant, K.R., Turbak, F.A., Brobst, S.A., Cohen, M.D. Intelligent information sharing systems. *Communications of the ACM*, 30(5), 1987, 390-402.
7. Jennings, A., Higuchi, H. A personal news service based on a user model neural network. *IEICE Transactions on Information and Systems*. E75-D(2) 198-210.
8. Desjardins G, Godin R. Combining relevance feedback and genetic algorithms in an Internet information personalization engine. *RIAO'2000 Conference Proceedings.* Vol.2, 2000, Paris, France.
9. Abidi, S.S.R., Han, C. Constraint Satisfaction Methods for Information Personalization. In *A. Tawfik, S. Goodwin (Eds.) Lecture Notes in Artificial Intelligence 3060, 17tth Canadian Conference. on Artificial Intelligence,* London, Ontario, 2004.
10. Abidi, S.S.R., Han, C. An Adaptive Hypermedia System for Information Customization via Content Adaptation. *IADIS International Journal of WWW/Internet,* 2(1), 2004, 79-94
11. Han, J.W., Kamber, M. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2000.
12. Tsang E, Foundations of constraint satisfaction. Academic Press, London, UK. 1993.
13. Barták, R. Constraint programming: In pursuit of the holy grail. *Proceedings of the Week of Doctoral Students (WDS99),* Part IV, MatFyz Press, Prague, 1999, 555-564.
14. Torrens, M., Faltings, B. SmartClients: Constraint satisfaction as a paradigm for scaleable intelligent information systems. *Workshop on Artificial Intelligence on Electronic Commerce, AAAI-99*, 1999, Florida, USA.
15. Padmanabhuni, S., You, J.H., Ghose, A. A framework for learning constraints. *Proc. of the PRICAI Workshop on Induction of Complex Representations,* August 1996.
16. O'Sullivan, B., Freuder, E.C.., O'Connell, S. Interactive Constraint Acquisition. In *Workshop on User-Interaction in Constraint Satisfaction, Seventh International Conference on Principles and Practice of Constraint Programming* - CP 2001, 2001.
17. Brin, S., Motwani, R., Silverstein, C. Beyond Market Baskets - Generalizing Association Rules to Correlations. In *Proceedings of the ACM SIGMOD*, 1997.
18. Freuder, E., Wallace, R. Partial Constraint Satisfaction. *Artificial Intelligence*, Vol. 58, 1992, 21-70.
19. Aarts, E., Lenstra, J.K., eds. *Local search in combinatorial optimization*. Princeton University Press, Princeton, NJ, 2003.
20. Meseguer, P., Bouhmala, N., Bouzoubaa, T., Irgens, M., Sanchez, M. Current Approaches for Solving Over-Constrained Problems. *Constraints*, Vol. 8, 2003, 9-39.