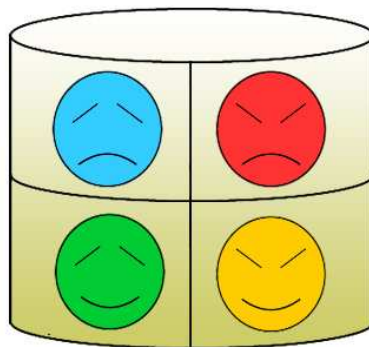


# **PersDB 2010**

4<sup>th</sup> International Workshop on  
Personalized Access, Profile Management, and  
Context Awareness in Databases



## **ELECTRONIC PROCEEDINGS**

In conjunction with VLDB 2010

September 13, 2010  
Singapore

## Foreword

Proliferation of database-driven web sites has brought upon a plethora of information access and dissemination applications. Monitoring and trading stock portfolios, news notification, weather tracking, and even simple search are just a few examples. In addition, recently, new applications have emerged that go beyond information access and dissemination and enhance creativity, information sharing, and collaboration among users providing richer interaction possibilities. Now, users cannot only access content but they can also generate, share and modify content (both theirs and others') freely, compose their applications, enhance their interface, etc. Web-based communities, wikis, social networks, mashups, folksonomies are some of the emerging new applications. In all these (classical and novel) applications, different notions of user information, such as preferences, community memberships and social interactions, and context information, such as a user's social network, location, time, and other features of a user's environment, are of paramount importance in order to improve and personalize user experience. In this context, new challenges emerge for user-centric, context-aware database systems for storing and managing different aspects of user and context information, for data management and computing taking into account personal, social and contextual information about users and for customizability of their behavior.

User-centric, personalized, socially-affected, and context-aware database systems represent a remarkable step towards user-centric applications that allow users to find, generate, share and modify content and services. It is imperative that people studying and working on the different components of a database system clarify their view of contextualization and personalization and describe the ways and degree to which these can be applied to their components of interest. We need a common understanding of the new challenges and we need to design new models, new algorithms, new databases to count for the user-centric requirements of emerging applications. The PersDB 2010 workshop aims at providing a forum for presentation of the latest research results, new technology developments, and new applications in the areas of personalized/socialized access, profile management, and context awareness in database systems.

The program committee consisted of 18 members and was chaired by Tiziana Catarci (Universita di Roma "La Sapienza", Italy) and Yannis Stavarakas (Institute for the Management of Information Systems, Greece). We would like to thank all the people who have supported and helped in the organization of this workshop: the authors and presenters of the papers for their interest and participation in PersDB 2010, the reviewers for their time and effort, and the organizers.

### Program Chairs

Tiziana Catarci (Universita di Roma "La Sapienza", Italy)

Yannis Stavarakas (Institute for the Management of Information Systems, Greece)

### Steering Committee

Tiziana Catarci (Universita di Roma "La Sapienza", Italy)

Yannis Ioannidis (University of Athens, Greece)

Georgia Koutrika (IBM Almaden Research Center, USA)

## Workshop Organization

### Steering Committee

Tiziana Catarci (Universita di Roma "La Sapienza", Italy)

Yannis Ioannidis (University of Athens, Greece)

Georgia Koutrika (IBM Almaden Research Center, USA)

### Program Committee Chairs

Tiziana Catarci (Universita di Roma "La Sapienza", Italy)

Yannis Stavarakas (Institute for the Management of Information Systems, Greece)

### Program Committee

Wolf Tilo Balke (University of Hannover, Germany)

Michael Benedikt (Oxford University, UK)

Paolo Ciaccia (University of Bologna, Italy)

Irene Fundulaki (ICS-FORTH, Greece)

Werner Kiessling (University of Augsburg, Germany)

Alexandros Labrinidis (University of Pittsburgh, USA)

Carlo Meghini (CNR, Italy)

Massimo Melucci (University of Padua, Italy)

Mohamed Mokbel (University of Minnesota, USA)

Moira Norrie (ETH-Zentrum, Switzerland)

Panagiotis Papadimitriou (Stanford University, USA)

Christos Papatheodorou (Ionian University, Greece)

Evi Pitoura (University of Ioannina, Greece)

Guillaume Raschia (Polytech' Nantes, France)

Letizia Tanca (Polytecnico di Milano, Italy)

Martin Theobald (Max Plank Institute, Germany)

Xi Zhang (University at Buffalo-SUNY, USA)

Xiaofang Zhou (The University of Queensland, Australia)

### PersDB 2010 Website

Anastasios Arvanitis (National Technical University of Athens, Greece)

## Table of Contents

### RESEARCH PAPERS

#### **A Benchmark for Context Data Management in Mobile ContextAware Applications**

Nadine Froehlich, University Basel  
Thorsten Moeller, University Basel  
Steven Rose, University Basel  
Heiko Schuldt, University Basel

#### **Combining Preference Relations: Completeness and Consistency**

Nicolas Spyratos, University of Paris-South  
Carlo Meghini, CNR-Italy

#### **Personalization through Query Explanation and Document Adaptation**

Anthony Ventresque, NTU-Singapore  
Sylvie Cazalens, Université de Nantes  
Thomas Cerqueus, Université de Nantes  
Philippe Lamarre, Université de Nantes  
Gabriella Pasi, Università di Milano Bicocca

#### **SQL QueRIE Recommendations: a query fragment-based approach**

Jayad Akbarnejad, San Jose State University  
Magdalini Eirinaki, San Jose State University  
Suju Koshy, San Jose State University  
Duc On, San Jose State University  
Neoklis Polyzotis, UC Santa Cruz

#### **Re-ranking Web Service Search Results Under Diverse User Preferences**

Dimitrios Skoutas, L3S Research Center  
Mohammad Alrifai, L3S Research Center  
Wolfgang Nejdl, L3S Research Center

#### **Adapting Generic Web Structures with Semantic Web Technologies: A Cognitive Approach**

Mario Belk, University of Cyprus  
Panagiotis Germanakos, University of Cyprus  
Nikos Tsianos, University of Athens  
Zacharias Lekkas, University of Athens  
Costas Mourlas, University of Athens  
George Samaras, University of Cyprus

# Combining Preference Relations: Completeness and Consistency

Nicolas Spyratos  
Université Paris-Sud  
Laboratoire de Recherche en Informatique  
Orsay Cedex, France  
spyratos@lri.fr

Carlo Meghini  
Consiglio Nazionale delle Ricerche  
Istituto della Scienza e delle Tecnologie della  
Informazione  
Pisa, Italy  
meghini@isti.cnr.it

## ABSTRACT

We introduce two criteria for judging “goodness” of the result when combining preference relations in information systems: completeness and consistency. Completeness requires that the result must be the union of all preference relations, while consistency requires that the result must be an acyclic relation. In other words, completeness requires that the result contain all pairs appearing in the preference relations, and only those pairs; while consistency requires that for every pair  $(x, y)$  in the result, it must be able to decide which of  $x$  and  $y$  is preferred to the other. Obviously, when combining preference relations, there is little hope for the result to satisfy both requirements. In this paper, we classify the various methods for combining preference relations, based on the degree to which the result satisfies completeness and consistency. Our results hold independently of the nature of preference relations (quantitative or qualitative); and also independently of the preference elicitation method (*i.e.* whether the preference relations are obtained by the system using query-log analysis or whether the user states preferences explicitly). Moreover, we assume no constraints whatsoever on the preference relations themselves (such as transitivity, strict ordering and the like).

## 1. INTRODUCTION

The problem considered in this paper is how to combine preferences to arrive at a consensus in the context of databases and information systems. Combining preferences to arrive at a consensus is a problem known in the literature as Social Choice Theory [20]. Variants of this problem, such as voting schemes, have been studied since the 18th century (by Condorcet and Borda). More recently, preferences have been used since the 50s in decision making for ranking alternative choices (see [18] for an extensive survey). In databases and information systems, however, the use of preferences started only in the late 90s and it is mainly concerned with the ranking of query answers [13, 7, 19, 1, 6, 4,

5, 12, 15, 14]. Indeed, with the explosion of the amount of information available today (*e.g.*, on the web) query results may be very large, and ranking these results according to users’ preferences is of great help to users.

In the area of information systems, user preferences are classified from different points of view, as follows.

In terms of their nature, preferences can be:

- *Quantitative* (or absolute), expressed by a number on a scale (thus capturing intensity of desire). For example, “I like BMW cars 80%” and “I like VW cars 70%”. Quantitative preferences are difficult to express by the casual user, but easy to compute by a machine from query logs.
- *Qualitative* (or relative), expressed by comparison. For example, “I like BMW more than VW”. Qualitative preferences express no intensity of desire; they are easy to express by the casual user and also easy to infer by a machine (from quantitative preferences).

In terms of their duration in time, preferences can be:

- *Long-term* preferences; these may be either discovered by the system (unobtrusively, from query logs) or declared explicitly by the user. In both cases long-term preferences are stored in the so-called user profile.
- *Short-term* preferences; these are expressed explicitly by the user, on-line, together with a query (in which case one usually talks about *preference queries*).

We note that the nature and the duration in time are orthogonal characteristics of preferences.

In this paper, we assume a set of objects  $O$  and we model a preference over  $O$  as a pair  $(o, o')$  meaning that  $o$  is preferred to  $o'$ ; moreover, we refer to a set of preferences over  $O$  as a *preference relation*.

We stress the fact that we model preference relations as binary relations over  $O$  *without* any particular constraint such as transitivity, strict ordering and the like. The reason for doing so is that there is no general consensus in the literature as to what properties a preference relation should satisfy. For instance, although transitivity is generally considered as a desirable property, in several situations preference relations are assumed to be non-transitive [9, 10, 11, 22].

Our approach considers only positive preference statements of the kind “ $x$  is preferred to  $y$ ”; in other words, we do not take into consideration indifference relations [18].

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, to post on servers or to redistribute to lists, requires a fee and/or special permission from the publisher, ACM.

VLDB '07, September 23-28, 2007, Vienna, Austria.  
Copyright 2007 VLDB Endowment, ACM 978-1-59593-649-3/07/09.

When  $O$  represents a set of alternative choices, one or more experts may be asked to express their opinion directly on the alternatives, thus leading to a set of preference relations over  $O$ . In other cases, the experts may be asked to express their opinion on one or more features of the alternatives, each one inducing a preference relation over  $O$ . In both cases, we end up with a set of preference relations over  $O$ , and the problem is how to combine them into a single preference relation incorporating as best as possible the opinions of all experts.

In the area of databases and information systems, there are two general methods for combining a set of preference relations: the Prioritized method and the Pareto method, in their restricted and unrestricted versions. Several studies in the literature use these methods, for example for defining preference queries and sky-line queries [3]. However, to our knowledge, no previous work has addressed the problem of how well the information content of the individual preferences is incorporated in the combined relation by these methods.

In this paper, we propose to evaluate Prioritized and Pareto, and their variants, by introducing two criteria: completeness and consistency; and we classify these methods based on the degree to which the combined relations satisfy the two criteria.

Our results hold independently of the nature of preference relations (*i.e.* whether they are qualitative or whether they have been inferred from quantitative rankings); and also independently of the preference elicitation method (*i.e.* whether the preference relations have been obtained by the system, using query-log analysis, or whether they have been stated by the user states, explicitly).

Section 2 gives the formal statement of the problem, as it appears in the area of information systems, also providing some rationale; section 3 introduces the classical methods for combining preference relations; section 4 analyzes how these methods behave with respect to the measures introduced earlier, namely consistency and completeness. Finally, Section 5 summarizes the results and draws some conclusions. Proofs are omitted for reasons of space.

## 2. STATEMENT OF THE PROBLEM AND RATIONALE

We say that a binary relation  $P$  is *complete* with respect to  $n > 1$  given binary relations  $P_1, \dots, P_n$  iff  $P$  is the union of  $P_1, \dots, P_n$ . Moreover, we say that  $P$  is *consistent* iff it is acyclic. We note that both completeness and consistency [8] can be tested efficiently.

Given a set of preference relations  $P_1, \dots, P_n$ , the problem that we consider is how to find a *combined* preference relation  $P$  that satisfies the following requirements:

1. *Completeness.* This property requires that the result should contain all pairs appearing in the preference relations, and only those pairs (*i.e.*, no preference expressed by the user is lost, and no extraneous preference is introduced in the result).
2. *Consistency.* This property requires that the result must be an acyclic relation; that is, for every pair  $(x, y)$  appearing in the result, it must be able to decide which of  $x$  and  $y$  is preferred to the other.

Clearly, when the result is complete, all expressed preferences are taken into account, but there may be contradictions among the individual preference relations, generating cycles in their union. On the other hand, consistency expresses absence of contradictions. So the presence of both completeness and consistency characterizes the optimal situation, where all preferences are taken into account and there is no contradiction.

We note that the presence of contradictions is natural, as they are the result of putting together preferences which either come from different users, independently, or come from the same user but address different aspects of the objects. As we mentioned earlier, such non-contradiction is expressed by the absence of cycles. Indeed, acyclicity allows the ranking of objects, as follows [21, 17]. Let  $P$  be an acyclic binary relation (viewed as a digraph), and define the rank of an object  $o$  as follows:

- if  $o$  is a root of  $P$  then  $rank(o) = 0$  (a root is a node with no incoming edge);
- else  $rank(o)$  is the length of a maximal path among all paths from a root of  $P$  to  $o$ .

The intuition behind this definition of rank is that the farther an object  $o$  is from the best objects (represented by the roots) the less preferred it is. We note that the computational complexity of computing ranks following this definition is linear in the size of the preference graph (understood as usual as the number of nodes plus the number of arcs). In reality, the size of the preference relations is quite small as users often express just a few preferences. Clearly, the definition of rank is sound only if  $P$  has at least one root, and this is guaranteed only when  $P$  is acyclic.

Now, let us denote by  $B_i$  the set of objects with rank  $i$ , and let  $m$  be the maximal path length among all paths starting from a root. Then it is easy to prove the following proposition:

**PROPOSITION 1.** *The sequence  $B_0, B_1, \dots, B_m$  defined above has the following properties:*

- $B_0, B_1, \dots, B_m$  is a partition of the set of objects appearing in  $P$ ;
- for each  $i = 0, 1, 2, \dots, m$ , there is no arc of  $P$  connecting two objects of  $B_i$ ;
- for each  $i = 1, 2, \dots, m$  and each object  $t \in B_i$  there is an object  $s \in B_{i-1}$  such that there exists an arc from  $s$  to  $t$  in  $P$ .

In general, when combining preference relations, there is little hope for the result to satisfy both completeness and consistency. Since ranking is important in information systems and acyclicity is a sufficient condition in order to do ranking, a reasonable approach to follow when combining preference relations is to satisfy acyclicity while *minimizing* the loss of completeness. In order to achieve this goal, we must select as the result  $P$  of combining the given preference relations a largest acyclic subset of the union. In other words, we must select  $P$  in such a way that there is no proper superset of  $P$  which is an acyclic subset of the union.

If we view preference relations as digraphs, finding such a  $P$  is equivalent to solving the maximum acyclic sub-graph problem, which is stated as follows: Given a digraph  $G =$

$(V, E)$ , find a maximum cardinality subset  $E' \subseteq E$  such that  $(V, E')$  is acyclic. This problem is known to be NP-hard [16].

Now, since there have been several proposals in the literature for combining preference relations, one wonders how these proposals relate to the maximum acyclic sub-graph problem. This is the question that we undertake in the rest of this paper.

More specifically, we show that the classical methods for combining preference relations provide a clever trade-off between maximality of the result and efficiency of computation. As these methods were introduced long before the theory of complexity was formulated, this is a rather surprising result that provides an *a posteriori* justification for the introduction of these methods.

### 3. CLASSICAL METHODS FOR COMBINING PREFERENCE RELATIONS

There are several well-known methods for combining a set of preference relations  $P_1, \dots, P_n$ , into a single preference relation, most notably, the so-called Prioritized and Pareto (and their variants). In order to formally define these methods, we use the following terminology from [2]. Let  $P$  be a preference relation:

- we use interchangeably the notations  $(x, y) \in P$  and  $xPy$ ; if  $(x, y) \notin P$ , we write  $x\bar{P}y$ ;
- $x$  and  $y$  are said to be *equivalent* with respect to  $P$ , written  $xP^\#y$ , iff both  $xPy$  and  $yPx$  hold;
- if  $x\bar{P}y$  and  $y\bar{P}x$ , we say that  $x$  and  $y$  are *incomparable* with respect to  $P$ , written  $xP^\#y$ ;
- finally, if  $xPy$  and  $y\bar{P}x$ ,  $x$  is said to be *strictly preferred* to  $y$  with respect to  $P$ , written  $xP^<y$ .

We first recall the definition of  $P_\cup$ , the union of the given preference relations  $P_1, \dots, P_n$ :

$$xP_\cup y \iff \exists i.(xP_i y)$$

A basic difference between Prioritized and Pareto is that the former, apart from the given preference relations  $P_1, \dots, P_n$ , requires some additional information in order to be applied, whereas the latter requires no additional information. Indeed, in the case of Prioritized, one assumes that the preference relations  $P_1, \dots, P_n$ , are ordered by a priority relation  $\prec$ , which is a strict partial order. We recall that a strict partial order is a binary relation which is irreflexive and transitive, and consequently asymmetric (asymmetric means that if  $a < b$  holds then  $b < a$  does not hold).

Prioritized and Pareto each come into two variants:

#### 1. Restricted Prioritized (RPR, for short).

This is the “classical” Prioritized rule, also used in [2]. Given a set of preference relations  $P_1, \dots, P_n$  strictly ordered by  $\prec$ , the RPR rule defines a binary relation  $Pr_r(P_1, \dots, P_n, \prec)$ , or simply  $Pr_r$  when there is no ambiguity, as follows:

$$xPr_r y \iff \forall i.(xP_i y \vee \exists j.(j \prec i \wedge xP_j^<y))$$

Notice that  $x$  and  $y$  are incomparable in the combined relation (*i.e.*,  $xPr_r^\#y$ ) if and only if  $x$  and  $y$  are incomparable on the preference relation with the highest priority for which they are not equivalent.

We note that the well-known lexicographic ordering is a special case of Restricted Prioritized. Indeed, the lexicographic ordering is a Prioritized ordering, in which:

- $O$  stands for a set of words of finite length over some finite alphabet  $A$ , where  $A$  is totally ordered by some strict total order  $<_A$  (*i.e.*, given any two distinct letters  $a$  and  $a'$ , either  $a <_A a'$  or  $a' <_A a$  but not both); the minimum element of  $<_A$  is the special character *blank*;
- we have as many preference relations  $P_i$  as the number of characters in the longest word in  $O$ ;
- the  $i$ -th preference relation captures preference based on the  $i$ -th letter of words (starting from the left), according to  $<_A$ ; so, *carlo* $P_1$ *nicolas* while *nicolas* $P_3$ *carlo*;
- the priority relation over the  $P_i$ 's is the natural total order  $<_N$  over the indices.

Under these assumptions, given two words  $w = a_1 a_2 \dots a_m$  and  $w' = a'_1 a'_2 \dots a'_n$ , the RPR rule becomes:

$$wPr_r w' \iff \forall i.(a_i <_A a'_i \vee \exists j.(j <_N i \wedge a_j <_A^<a'_j))$$

Now, since  $<_A$  is a strict total order,  $<_A = <_A^<$ , therefore the rule becomes:

$$wPr_r w' \iff \forall i.(a_i <_A a'_i \vee \exists j.(j < i \wedge a_j <_A a'_j))$$

The last rule is clearly the rule used to order words in a dictionary.

#### 2. Unrestricted Prioritized (UPR, for short).

This method extends RPR by allowing the combined preference to hold even in presence of incomparability in some preference relation, as long as there exists comparability in some other relation. The UPR rule defines a binary relation  $Pr_u$  as follows:

$$xPr_u y \iff \forall i.(xP_i y \vee (xP_i^\#y \wedge \exists k.(xP_k y)) \vee \exists j.(j \prec i \wedge xP_j^<y))$$

Notice that  $x$  and  $y$  are incomparable if and only if  $x$  and  $y$  are incomparable in *all* given preference relations.

#### 3. Restricted Pareto (RPA, for short).

In this method,  $x$  is preferred to  $y$  in the combined relation if and only if for every  $i$ ,  $x$  is preferred to  $y$ , and for at least one  $P_i$   $x$  is strictly preferred to  $y$ . Formally, the RPA rule defines a binary relation  $Pa_r$  as follows:

$$xPa_r y \iff \forall i.(xP_i y) \wedge \exists j.(xP_j^<y)$$

#### 4. Unrestricted Pareto (UPA, for short).

In this method,  $x$  is preferred to  $y$  if and only if for at least one  $P_i$   $x$  is strictly preferred to  $y$ , and for no  $j$   $y$  is strictly preferred to  $x$  (in all other preferences  $x$  and  $y$  can be comparable in the same direction, incomparable or equivalent). The UPA rule defines a binary relation  $Pa_u$  as follows:

$$xPa_u y \iff \forall i.(y\bar{P}_i^<x) \wedge \exists j.(xP_j^<y)$$

Based on the definitions given so far we can state the following proposition.

PROPOSITION 2. Let  $P_1, \dots, P_n$ , be  $n$  preference relations and let  $\prec$  be any strict ordering on them. Then, the following hold:

1.  $Pa_r \subseteq Pr_r \subseteq Pr_u \subseteq P_U$
2.  $Pa_r \subseteq Pa_u \subseteq Pr_u \subseteq P_U$
3.  $Pa_u$  and  $Pr_r$  are incomparable with respect to set-containment.

#### 4. COMPLETENESS AND CONSISTENCY OF CLASSICAL METHODS

In order to characterize the classical methods with respect to completeness and consistency, we consider three cases as follows:

**Case 1** the union  $P_U$  is acyclic.

In this case, consistency is guaranteed and we show that completeness holds only for the unrestricted methods.

**Case 2** the union  $P_U$  is cyclic but each individual preference relation is acyclic.

In this case, we show that the restricted methods lead to an acyclic result, but they are incomplete in the sense that they may not produce a maximum acyclic sub-graph; on the other hand, the unrestricted methods may produce a cyclic result.

**Case 3** one or more individual preference relations are cyclic.

In this case, all methods may produce a cyclic result.

##### 4.1 Acyclic union

It follows from proposition 2 that whenever the union of the preference relations is acyclic, so are all the variants of Prioritized and Pareto defined above. In this case, then, consistency is satisfied. It remains to be seen whether completeness is satisfied as well.

The next lemma states a consequence of the acyclicity of the union that will be used in what follows.

LEMMA 1. Let  $P_1, \dots, P_n$ , be preference relations whose union  $P_U$  is acyclic. If  $xP_U y$  then for some  $i$  we have that  $xP_i^<y$ , and for all  $k \neq i$  we have that either  $xP_k y$  or  $xP_k^\# y$ .

We can now prove the following.

PROPOSITION 3. Let  $P_1, \dots, P_n$  be preference relations whose union  $P_U$  is acyclic. Then,

1.  $P_U = Pa_u$
2.  $P_U = Pr_u$  for any strict order  $\prec$  of the preference relations.

Notice that in proving that  $Pr_u \subseteq P_U$ , the acyclicity of the union is not required.

Concerning  $Pr_r$  and  $Pa_r$ , we have that in some cases  $P_U$  is acyclic and that  $Pr_r \subset P_U$  and  $Pa_r \subset P_U$ . In proof, let us consider the example shown in Figure 1; using RPR to combine  $P_1$  and  $P_2$  leads to the incomparability of 3 and 4 (if  $P_1 \prec P_2$ ) or the incomparability of 1 and 2 (if  $P_2 \prec P_1$ ),

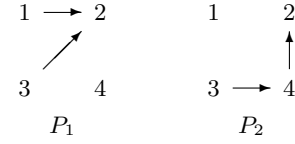


Figure 1: Preference relations

while both these pairs are in  $P_U = P_1 \cup P_2$ . Analogously, it can be seen that  $(3, 4) \in Pa_r^\#$  and  $(1, 2) \in Pa_r^\#$ .

We can then conclude that when the union of the given preference relations is acyclic, both RPR and RPA are incomplete, while UPR and UPA are “optimal”, *i.e.* both complete and consistent.

##### 4.2 Cyclic union with acyclic preference relations

We now consider the case in which the union of the preference relations is cyclic while each individual preference relation is acyclic. In this case, we are interested to know (a) whether Prioritized and Pareto produce an acyclic combined preference relation, and, if yes, (b) how close to the union  $P_U$  they are.

In order to answer these questions, we first derive necessary and sufficient conditions for the acyclicity of the result in each of the classical methods. We then apply these conditions to answer the above questions.

We recall that a cycle in a binary relation  $P$  is a sequence of objects  $C = (o_0, o_1, \dots, o_k)$  with  $k \geq 2$ , such that

- $o_0 = o_k$ ,
- $o_{i-1}Po_i$  for each  $i = 1, \dots, k$ , and
- there is no repetition in  $o_0, o_1, \dots, o_{k-1}$ .

PROPOSITION 4. Let  $P_1, \dots, P_n$ , be  $n$  preference relations and let  $\prec$  be any strict ordering on them. Then, the Restricted Prioritized  $Pr_r$  is acyclic iff for each cycle  $C = (o_0, \dots, o_k)$  in  $P_U$  there exists an arc  $(o_{u-1}, o_u)$  such that for some preference relation  $P_i$ ,  $o_{u-1}P_i o_u$  and for each  $j$  such that  $o_{u-1}P_j o_u$ , either  $o_u P_j o_{u-1}$  or  $i \prec j$ .

The last Proposition states the condition under which an undesired preference  $(o, o')$  (*i.e.*, any one of the preferences found in a cycle in  $P_U$ ) does not end up in  $Pr_r$ . The condition is derived by combining two facts:  $oP_i o'$  and  $o'P_r o$  (the latter in turn obtained by negating the RPR rule). By a similar technique, we obtain the analogous conditions for the other considered variants of Prioritized and Pareto. The proofs of the corresponding Propositions are omitted, as they are very similar to that of the previous Proposition.

PROPOSITION 5. Let  $P_1, \dots, P_n$ , be  $n$  preference relations and let  $\prec$  be any strict ordering on them. Then, the Unrestricted Prioritized  $Pr_u$  is acyclic iff for each cycle  $C = (o_0, \dots, o_k)$  in  $P_U$  there exists one arc  $(o_{u-1}, o_u)$  such that for some preference relation  $P_i$ ,  $o_u P_i^<o_{u-1}$  and for each  $j$  such that  $o_{u-1}P_j o_u$ , either  $o_u P_j o_{u-1}$  or  $i \prec j$ .

PROPOSITION 6. Let  $P_1, \dots, P_n$ , be  $n$  preference relations. Then, the Restricted Pareto preference relation  $Pa_r$  is acyclic iff for each cycle  $C = (o_0, \dots, o_k)$  in  $P_U$  there exists one arc  $(o_{u-1}, o_u)$  such that either  $o_{u-1}P_i o_u$  for some preference relation  $P_i$ , or  $o_u P_j o_{u-1}$  for all preference relations  $P_j$ .



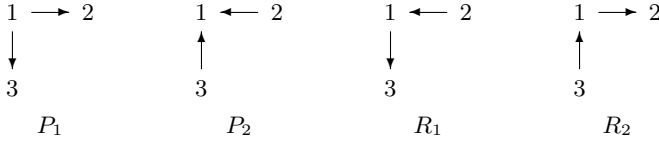


Figure 2: Prioritized combinations

PROPOSITION 7. Let  $P_1, \dots, P_n$ , be  $n$  preference relations. Then, the Unrestricted Pareto  $Pa_u$  is acyclic iff for each cycle  $C = (o_0, \dots, o_k)$  in  $P_U$  there exists one arc  $(o_{u-1}, o_u)$  such that  $o_u P_i^< o_{u-1}$  for some preference relation  $P_i$  or  $o_{u-1} P_j o_u$  implies  $o_u P_j o_{u-1}$  for all  $j$ .

Having established under which condition each of the considered methods produces a cyclic relation, we now examine whether this condition can ever occur.

PROPOSITION 8. The Restricted Prioritized  $Pr_r$  of  $n$  acyclic preference relations  $P_1, \dots, P_n$  is acyclic for any strict order  $\prec$  on  $P_1, \dots, P_n$ .

The last Proposition says that when each one of the given preference relations is acyclic, the necessary and sufficient conditions for the acyclicity of  $Pr_r$ , established by Proposition 4 are always met. This can be verified by considering that under the circumstances, every cycle in  $P_U$  involves at least two preference relations, one of which has necessarily a lower priority than the other; thus, any arc  $(x, y)$  in the cycle coming from a preference relation with a lower priority can play the role of  $(o_{u-1}, o_u)$  in Proposition 4.

The last Proposition shows that in presence of a cyclic union, the RPR rule produces an acyclic relation. However, the rule turns out to be too restrictive in that it may exclude from the result preferences that are unproblematic (*i.e.*, preferences which do not produce any cycle if included in the result). In other words, RPR may not produce a maximum acyclic sub-graph of  $P_U$ .

Consider for instance the following preference relations:  $P_1 = \{(2, 3)\}$ ,  $P_2 = \{(1, 2)\}$  and  $P_3 = \{(3, 2)\}$ , together with the following strict order:  $P_1 \prec P_2 \prec P_3$ . In this case, we have  $Pr_r = \{(2, 3)\}$ , which is a non-maximum acyclic sub-graph of  $P_U$ , since it misses the preference  $(1, 2)$ .

Furthermore, there may be maximum acyclic sub-graphs of  $P_U$  that RPR cannot define under any  $\prec$ . The example presented in Figure 2 is a case in point: we have  $Pr_r = P_1$  (if  $P_1 \prec P_2$ ) or  $Pr_r = P_2$  (if  $P_2 \prec P_1$ ). But for no ordering of the preference relations we can have  $Pr_r = R_1$  or  $Pr_r = R_2$  (also shown in Figure 2).

Let us now consider  $Pr_u$ . It is easy to construct an example that falsifies the condition of Proposition 5 (*i.e.*, a cyclic  $P_U$  leading to a cyclic  $Pr_u$ ): Assume  $P_1 \prec P_2$  and  $P_1 = \{(1, 2)\}$  whereas  $P_2 = \{(2, 3), (3, 1)\}$ . In this case, none of the arcs making up the cycle  $(1, 2, 3, 1)$  in  $P_U$  has a reverse arc as required by Proposition 5. In fact,  $Pr_u = P_U$  and so we have a cyclic  $Pr_u$ .

For Pareto, the situation is identical.

COROLLARY 1. The Restricted Pareto  $Pa_r$  of  $n$  acyclic preference relations  $P_1, \dots, P_n$  is acyclic.

As for Proposition 8, the acyclicity of each preference relation  $P_i$  implies that the condition established by Proposition 6 are always met. In particular, each cycle in  $P_U$  comes

from preferences belonging to different relations; then, any arc  $(x, y)$  in the cycle is missed at least by one  $P_i$ , *i.e.*  $x P_i y$ , and as such it plays the role of  $(o_{u-1}, o_u)$  in Proposition 6.

Since  $Pa_r$  is a subset of  $Pr_r$  (Proposition 2), we have that the above remarks related to the completeness of  $Pr_r$  carry over to  $Pa_r$ .

Finally, the previous example can be used also to show that a cyclic  $P_U$  may lead to a cyclic  $Pa_u$ .

We may then conclude that when each preference relation is acyclic but their union is cyclic, both  $Pr_r$  and  $Pa_r$  are acyclic but incomplete. On the other hand, their unrestricted versions  $Pr_u$  and  $Pa_u$  may be cyclic.

### 4.3 Cyclic preference relations

In considering cyclic preference relations, we distinguish between two cases:

- the cycle involves only two objects.

We call these objects *equivalent* in the sense defined earlier.

- the cycle involves at least three objects.

In the classical methods, object equivalence arises only in very special circumstances or not at all, as shown by the following corollary (that follows from Propositions 4 to 7):

COROLLARY 2. Let  $P_1, \dots, P_n$ , be  $n$  preference relations and let  $\prec$  be any strict ordering on them. Then, for any two objects  $x, y$

- $x Pr_r^= y$  iff  $x P_i^= y$  for all  $i = 1, \dots, n$ , and

- $x Pr_u^= y$  iff the following hold:

1. for some  $i$  we have  $x P_i^= y$ , and
2. for all  $j = 1, \dots, n$  we have either  $x P_j^= y$  or  $x P_j^# y$ .

Moreover, both  $Pa_r$  and  $Pa_u$  are asymmetric.

Thus, two objects are equivalent with respect to Restricted Prioritized if they are equivalent with respect to each preference relation. In this case they are incomparable for both versions of Pareto. For Unrestricted Prioritized, two objects are equivalent if they are comparable on some dimension but in no dimension one of the two is strictly preferred to the other. Pareto rules out object equivalence by being asymmetric in both its variants.

For cycles involving three or more objects, the situation is much worse. If some of the  $P_1, \dots, P_n$  contain such a cycle, then all variants of Prioritized and Pareto may produce a cyclic result.

The following example shows that this is indeed the case for RPA. Suppose  $P_1 = P_2 = \{(1, 2), (2, 3), (3, 1)\}$ . Then, it is easy to see that  $Pa_r = P_1$ . Notice that if we assume transitivity of the  $P_i$ 's, then for all objects  $x$  and  $y$  involved in a cycle, we have  $x P_i^= y$  and therefore  $(x, y) \notin Pa_r$ . In the example, transitivity of the  $P_i$ 's implies that  $P_i^< = \emptyset$  for all  $i$  and consequently  $Pa_r = \emptyset$ .

Since  $Pa_r$  is the smallest relation among those produced by the classical methods, we have that all methods can produce a cyclic result.

## 5. CONCLUDING REMARKS

From an information system perspective, there are two basic requirements that must be satisfied when combining preference relations into a single preference relation: completeness and consistency. Completeness means that the result must be the union of all preference relations, while consistency means that the result must be an acyclic relation.

However, it is rarely possible to satisfy these two requirements simultaneously, due to the fact that the union of the given preference relations may be cyclic. Given the importance of acyclicity in ranking the objects, we have argued that a reasonable compromise is to aim at a maximum acyclic sub-graph of the union of the given relations.

We have analyzed two classical methods for combining preference relations, Prioritized and Pareto, each in two variants: restricted and unrestricted. We have shown that all four methods are inadequate with respect to the above requirements, as they may produce a result that is either incomplete or cyclic. In particular,

- In the (fully unproblematic) case when the union of the given preference relations is acyclic, both restricted approaches are incomplete, while the unrestricted ones are optimal (*i.e.*, complete and acyclic).
- When each preference relation is acyclic but their union is cyclic, both restricted approaches produce an acyclic result, which however loses more preferences than needed to obtain acyclicity; under the same circumstances, the unrestricted methods may produce a cyclic result.
- Finally, when one or more preference relations have cycles involving at least three objects, all four methods may produce a cyclic result.

Thus, all four classical methods are for various reasons unsatisfactory. However, if we look at the problem from a purely computational perspective, we are faced with the fact that computing a maximum acyclic sub-graph of a given graph is known to be NP-hard. This fact sheds a different light on the classical approaches.

In fact we may conclude that both Prioritized and Pareto trade off efficiency to optimality (unless  $P=NP$ ). In their unrestricted variants, both methods achieve optimality only when it is computationally easy to do so, (*i.e.*, when the union is acyclic). In all other cases, they retain efficiency while losing optimality.

## 6. REFERENCES

- [1] R. Agrawal and E. L. Wimmers. A framework for expressing and combining preferences. In *Proc. of ACM SIGMOD*, pages 297–306, Dallas, USA, 2000.
- [2] H. Andreka, M. Ryan, and P.-Y. Schobbens. Operators and laws for combining preference relations. *Journal of Logic and Computation*, 12(1):13–53, 2002.
- [3] R.I. Brafman and C. Domshlak. Preference handling - an introductory tutorial. *AI Magazine*, 30(1), 2009.
- [4] J. Chomicki. Querying with intrinsic preferences. In *Proceedings of the 8th International Conference on EDBT*, pages 34–51, Prague, Czech Rep., 2002.
- [5] J. Chomicki. Semantic optimization of preference queries. In *Proc. of 1st Int. Sym. on Appl. of Constraint Databases*, number 3074 in LNCS, 2004.
- [6] J. Chomicki. Iterative modification and incremental evaluation of preference queries. In *Proc. of FoIKS*, pages 63–82, 2006.
- [7] Jan Chomicki. Preference formulas in relational queries. *ACM Trans. Database Systems*, 28(4):427–466, 2003.
- [8] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein. *Introduction to algorithms*. The MIT Press, 2nd edition, 2001.
- [9] Peter C. Fishburn. Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty*, 4(2):113–124, April 1991.
- [10] Oswald Huber. Nontransitive multidimensional preferences: Theoretical analysis of a model. *Theory and Decision*, 19:147–165, January 1979.
- [11] J. Kacprzyk and M. Roubens, editors. *Non-conventional preference relations in decision making*. Springer-Verlag, 1988.
- [12] W. Kiessling and G. Köstler. Preference SQL design, implementation, experiences. In *In Proceedings of 28th International Conference on Very Large Data Bases*, pages 990–1001, Hong Kong, China, 2002.
- [13] Werner Kiessling. Foundations of preferences in database systems. In *Proc. of VLDB*, pages 311–322, 2002.
- [14] Georgia Koutrika and Yannis E. Ioannidis. Constrained optimalities in query personalization. In *Proc. of ACM SIGMOD*, pages 73–84, 2005.
- [15] Georgia Koutrika and Yannis E. Ioannidis. Personalized queries under a generalized preference model. In *Proc. of ICDE*, pages 841–852, 2005.
- [16] Alantha Newman. Approximating the maximum acyclic subgraph. Master’s thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 2000.
- [17] E.M. Nguer and N. Spyrtos. A user-friendly interface for evaluating preference queries over tabular data. In *Proc. of SIGDOC08, the 26th ACM International Conference on Design of Communication*, Lisbon, Portugal, Sept 2008.
- [18] M. Öztürk, A. Tsoukiàs, and Ph. Vincke. Preference modelling. In J. Figueira M. Ehrgott, S. Greco, editor, *Multiple Criteria Decision Analysis: State of the Art Surveys*, pages 27–73. Springer Verlag, 2005.
- [19] P.Georgiadis, I.Kapantaidakis, M.Nguer, N.Spyrtos, and V.Christophides. Efficient rewriting algorithms for preference queries. In *Proc. of ICDE08, the 24th International Conference on Data Engineering*, Cancun, Mexico, April 2008.
- [20] Amartya Sen. Social choice theory. In K. J. Arrow and M.D. Intriligator, editors, *Handbook of Mathematical Economics*, volume 3 of *Handbook of Mathematical Economics*, chapter 22, pages 1073–1181. Elsevier, September 2005.
- [21] N. Spyrtos and V. Christophides. Querying with preferences in a digital library. In *Dagstuhl Seminar (N 05182), Federation over the Web*, number 3847 in LNAI, May 2005.
- [22] Tversky. Intransitivity of preferences. *Psychological review*, 76(1), 1969.