Combining Preference Relations: Completeness and Consistency

Nicolas Spyratos[1] Carlo Meghini[2]

[1]UP Sud - LRI, Paris
[2]CNR - ISTI, Pisa

PersDB 2010: Fourth VLDB Workshop on Personalized Access, Profile Management, and Context Awareness in Databases
Singapore, Sept. 13, 2010

Outline

Introduction

We consider the problem of combining preferences to arrive at a consensus, in the context of databases and information systems.

Well-known problem:

- Social Choice Theory
- voting schemes (since the 18th century)
- decision making for ranking alternative choices

In databases and information systems: ranking of query answers.

Types of preferences

- *Quantitative* preferences, expressed by a number on a scale.
  I like BMW 70%
  Difficult to express by the casual user, but easy to compute by a machine (*e.g.*, from query logs).

- *Qualitative* preferences, expressed by comparison.
  I like BMW more than VW
  Easy to express by the casual user and also easy to infer by a machine (*e.g.*, from quantitative preferences).

We focus on qualitative preferences.

O : a set of objects $(o, o', o_1, \ldots, x, y, z, \ldots)$
(o, o') : a preference, meaning that o is preferred to o'
preference relation: a finite set of preferences $(P, P_1, \ldots)$

No constraint such as transitivity, strict ordering and the like.

Only positive preference statements, no indifference relations

A preference relation may be:

- expressed directly on objects by an expert.
  Many experts $\longrightarrow$ many preference relations
- induced by preferences over features of objects.
  Many features $\longrightarrow$ many preference relations

How to combine many preference relations into a single preference relation incorporating *as best as possible* the opinions of all experts.

In databases and information systems, two methods for combining preference relations:

- Prioritized
- Pareto

in their restricted and unrestricted versions.

Several studies in the literature use these methods, for example for defining preference queries and sky-line queries.

How well the information content of the individual preferences is incorporated in the combined relation?

1. We define two criteria for assessing the goodness of a combination operator: *completeness* and *consistency*
2. we evaluate Prioritized and Pareto based on the degree to which they satisfy these two criteria.

The problem

P : a binary relation
$P_1, \ldots, P_n$: given binary relations, $n > 1$

P is *complete* with respect to $P_1, \ldots, P_n$ iff P is the union of $P_1, \ldots, P_n$.

P is *consistent* iff it is acyclic.

If $P_1, \ldots, P_n$ are preference relations and P is their combined preference relation, completeness and consistency of P are desirable.

1. Completeness requires that no preference expressed by the user is lost, and no extraneous preference is introduced in the result.
2. Consistency requires that 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. This allows to rank the objects (*e.g.*, by topological sorting).

There is little hope for complete and consistent combined preferences.

Since ranking is important and acyclicity is a sufficient condition in order to do ranking, a reasonable approach is:

- to satisfy acyclicity
- while *minimizing the loss* of completeness.

P must be a largest acyclic subset of the union (no proper superset of P is an acyclic subset of the union).

Finding P requires solving the maximum acyclic sub-graph problem:

> *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.

On the other hand, classical proposals for combining preference relations are all efficiently computable.

Why?

In other words,

Which of the two criteria do they relax? completeness, consistency, or both?

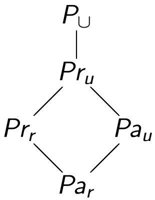
Notation

Let P be a preference relation:

- $x\overline{P}y$ means $(x, y) \notin P$
- $xP^{\equiv}y$ (x and y are *equivalent* with respect to P) iff both xPy and yPx hold
- $xP^{\#}y$ (x and y are *incomparable* with respect to P) iff $x\overline{P}y$ and $y\overline{P}x$
- $xP^{<}y$ (x is *strictly preferred* to y with respect to P, written) iff xPy and $y\overline{P}x$
- $xP_{\cup}y$ iff xP_iy for some i
- $\prec$ is a strict partial order over $P_1, \ldots, P_n$ (irreflexive and transitive, and consequently asymmetric).

Classical methods

$$
\begin{aligned}
xP_{\cup}y &\iff \exists i.(xP_i y)\\
xPr_r y &\iff \forall i.(xP_i y \ \vee\ \exists j.(j \prec i \wedge xP_j^{<} y))\\
xPr_u y &\iff \forall i.(xP_i y \ \vee\ (xP_i^{\#} y \wedge \exists k.(xP_k y)) \ \vee\ \exists j.(j \prec i \wedge xP_j^{<} y))\\
xPa_r y &\iff \forall i.(xP_i y) \ \wedge\ \exists j.(xP_j^{<} y)\\
xPa_u y &\iff \forall i.(y\overline{P_i^{<}}x) \ \wedge\ \exists j.(xP_j^{<} y)
\end{aligned}
$$



Analysis

We consider three cases:

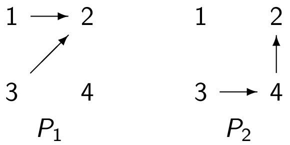
Case 1 the union $P_\cup$ is acyclic.
This obviously imples that all individual preference relations are acyclic.

Case 2 the union $P_\cup$ is cyclic but each individual preference relation is acyclic.
In this case the cycle is generated by arcs contributed by different preferences (*i.e.*, every expert is consistent but different experts contradict themselves).

Case 3 one or more individual preference relations are cyclic.
Contradictions arise within the given preferences.

Acyclic union (hence acyclic preference relations)

- Consistency is guaranteed by all methods, which produce subsets of the union, hence acyclic relations
- Completeness holds only for the unrestricted methods.



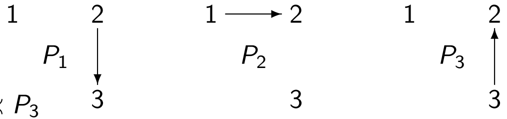
- if $P_1 \prec P_2$, then $(3, 4) \notin Pr_r$
- if $P_2 \prec P_1$, then $(1, 2) \notin Pr_r$

Analogously, it can be seen that $(3, 4) \notin Pa_r$ and $(1, 2) \notin Pa_r$.

Cyclic union, acyclic preference relations

If the union $P_\cup$ is cyclic but each individual preference relation is acyclic:

- 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.



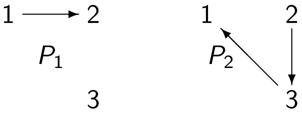
$P_1 \prec P_2 \prec P_3$

$(1,2) \notin Pr_r$, which is therefore not a maximum acyclic sub-graph of $P_\cup$.

For Pa_r the situation is the same.

Cyclic union, acyclic preference relations (cont.)

- The unrestricted methods may produce a cyclic result.



$P_1 \prec P_2$

$Pr_u = Pr_r = P_\cup$ so we have a cycle in each combined unrestricted relation

Cyclic preference relations (hence cyclic union)

We distinguish:

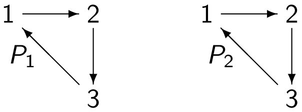
- cycles of length 1, object *equivalence*
- cycles of length at least 2

Object equivalence can be reproduced by Prioritized:

- Two objects are equivalent with respect to Restricted Prioritized if they are equivalent with respect to each preference relation.
- 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.

Cycles of length at least 2



In this case, $Pa_r = P_1 = P_2$ hence we have a cyclic combined preference relation.

Since Pa_r is the smallest relation among those produced by the classical methods, we have that if one or more individual preference relations are cyclic, then all methods may produce a cyclic result.

Conclusions

We propose two basic requirements as a measure of the adequacy of a combined preference relation: completeness and consistency.

When it is not possible to satisfy both these requirements, a reasonable compromise is to aim at a maximum acyclic sub-graph of the union of the given relations.

In the light of these criteria, we have analyzed two classical methods: 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.

However, 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 loosing optimality.