International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Approximations of arbitrary relations by partial orders

Ryszard Janicki

Department of Computing and Software, McMaster University, Hamilton L8S 4K1, Canada

A R T I C L E   I N F O

Article history:
Received 7 October 2017
Received in revised form 9 February 2018
Accepted 27 April 2018
Available online xxxx

Keywords:
???

A B S T R A C T

The problem of optimal quantitative approximation of an arbitrary binary relation by a partial order is discussed and some solutions are provided. It is shown that even for a very simple quantitative measure the problem is NP-hard. Some quantitative metrics are also applied for known property-driven approximations by partial orders. Some relationship to Rough Sets is discussed.

1. Introduction

A motivation for this kind of work has been clearly described in [13]:

"Consider the following problem: we have a set of data that have been obtained in an empirical manner. From the nature of the problem we know that the set should be partially ordered, but because the data are empirical it is not. In a general case, this relation may be arbitrary. What is the 'best' partially ordered approximation of an arbitrary relation and how this approximation can be computed?"

Areas of immediate applications of any 'best' partial order approximation algorithm include group ranking, social choice, pairwise comparisons based non-numerical ranking, analysis of subjective judgments, etc. [8,12,16,18].

Defining what is the 'best' partially ordered approximation is itself a problematic task. It could be approached in at least two ways.

One approach is just to propose some similarity metrics for binary relations and then just choose a partial order that is closest to a given arbitrary relation. This is the main subject of this paper. The first question is how these similarity metrics should be defined. Should we look for some generic similarity measure between arbitrary relations, or should we take into account that one of the relations is always a partial order and include this fact into the definition of similarity? Partial ordering means acyclity and transitivity, should our similarity measures also make this distinction? From the application point of view, the roles of acyclity and transitivity are different. Lack of transitivity may not be an error at all, it could just be a fact that a given data set is of minimum (or optimal, sufficient) size (as Hasse diagrams, directed acyclic graphs or dags [25], or direct causality graphs [21], etc.). Acyclity on the other hand is usually an indication of some errors. Moreover, the relation that we are going to approximate is not absolutely arbitrary. It represents, with perhaps some errors or incompleteness, some real data or phenomena. If its partially ordered approximation is chosen only on the basis of some numerical calculations, some structural properties of the original relation might be lost or wrongly changed.

E-mail address: janicki@mcmaster.ca.

https://doi.org/10.1016/j.ijar.2018.04.012
0888-613X/

The other orthogonal approach is not to use quantitative metrics at all. This approach is called *property-driven* and is based on the assumption that partial order approximations must satisfy certain properties. It stems from the 1895 paper by Schröder [26]. As partial orders, the approximations must be acyclic and transitive, but they also should satisfy some other properties. In [13] property-driven partial order approximations of an arbitrary binary relation were provided and discussed in both the classical algebraic model and the Rough Sets settings [24]. No quantitative metrics were used in [13].

In this paper we propose two simple metrics for measuring similarity and difference between relations, and a definition of *optimal* approximation. We also provide some justification of both metrics and the definition. One metric is just a simple adaptation of a metric used for sets, the other is a special modification designed specially for relations. We show that, at least for approximations by partial orders, both metrics lead to identical results.

In [13] and [14], a special attention is paid to two partially ordered approximations of R, denoted by $(R^\bullet)^+$ and $(R^+)^\bullet$ for a given relation R. Using graph terminology, $R^\bullet$ is derived from R by erasing all arcs from all strongly connected components (or equivalently, removing all arcs from all cycles). The relation R^+ is a transitive closure of R. The relation $(R^+)^\bullet$ is a classical approximation, first proposed by Schröder in 1895 [26], which is often regarded as 'the' partially ordered approximation. We will show that with respect to our metrics, $(R^\bullet)^+$ is a better approximation of R than the well known Schröder's $(R^+)^\bullet$.

We will also show that finding quantitative optimal approximation, with respect to simple metrics proposed in this paper, is NP-hard.

Finally, we will argue that while an arbitrary quantitative optimal approximation (with any reasonable metrics) might somehow be inconsistent with property-driven approximations of [13,14], our model satisfies most properties required from property-driven approximations.

We will also show how the presented model relates to the Rough Sets approach for specialized relational approximations.

This paper is a substantially extended, revised and corrected version of the conference paper [15].

The paper is organized as follows. In Section 2 we recall the basic notions of the theory of relations, directed graphs and partial orders. The basic concepts of similarity and distance for relations that are used in this paper, are presented in Section 3. The problems encountered when trying to define the concept of optimal approximation are discussed in Section 4, while Section 5 is devoted to the property-driven partial orders approximations of [13,14]. Quantitative properties of property-driven partial order approximations $(R^\bullet)^+$ and $(R^+)^\bullet$ are presented in Section 6. Some intuitions and properties that led to our concept of optimal partial order approximation are analyzed in Section 7. The main result of this paper, namely introduction and discussion of partial order approximations based on absolute similarity and distance, are presented in Section 8. Another version of a distance for relations is proposed and its properties are discussed in Section 9. In Section 10 the main results of the paper are presented in Rough Sets setting, and Section 11 contains final comments.

2. Relations, directed graphs and partial orders

In this section we recall some fairly known concepts and results that will be used later in this paper [3,7,25].

Let X be a set. *We assume all sets considered in this paper are finite.* Note that every relation $R \subseteq X \times X$ can be interpreted as a *directed graph* $G_R = (V, E)$ where $V = X$ is the set of vertices and $E = R$ is the set of edges (cf. [3]).

A relation $< \in X \times X$ is a (*sharp*) *partial order* if it is irreflexive and transitive, i.e. if $\neg(a < a)$ and $a < b < c \implies a < c$, for all $a, b, c \in X$.

We write $a \sim_< b$ if $\neg(a < b) \wedge \neg(b < a)$, that is if a and b are either *distinctly incomparable* (w.r.t. $<$) or *identical* elements. We also write

$$a \equiv_< b \iff (\{x \mid a < x\} = \{x \mid b < x\} \ \wedge \ \{x \mid x < a\} = \{x \mid x < b\}).$$

The relation $\equiv_<$ is an *equivalence relation* (i.e. it is reflexive, symmetric and transitive) and it is called *the equivalence with respect to* $<$, since if $a \equiv_< b$, there is nothing in $<$ that can distinguish between a and b (cf. [7]). We always have $a \equiv_< b \implies a \sim_< b$.

- Let $\mathbb{PO}(X)$ denote the set on all partial orders included in $X \times X$.

For every relation $R \subseteq X \times X$, we define $R^0 = Id_X = \{(a,a) \mid a \in X\}$, the identity relation on X, and $R^{i+1} = R^i R$ for all $i \geq 0$. Furthermore the relation $R^+ = \bigcup_{i=1}^{\infty} R^i$ is called the *transitive closure* of R, the relation $R^{-1} = \{(b,a) \mid (a,b) \in R\}$ is called the *inverse* of R, and a relation R is *acyclic* if and only if $\neg x R^+ x$ for all $x \in X$. In graph terminology, if R is acyclic then G_R is DAG (*Directed Acyclic Graph*), while if for all $x \in X$ we have xR^+x then the graph G_R is *strongly connected*. Also for a given relation R and $a \in X$, we define: $aR = \{x \mid aRx\}$ and $Ra = \{x \mid xRa\}$.

For every relation R we can define the relations R^{cyc}, $R^\bullet$ and $\equiv_R$ as follows

- $aR^{cyc}b \iff aR^+b \wedge bR^+a$,
- $aR^\bullet b \iff aRb \wedge \neg(aR^{cyc}b)$, i.e. $R^\bullet = R \setminus R^{cyc}$,
- $a \equiv_R b \iff aR = bR \wedge Ra = Rb$.

Please cite this article in press as: R. Janicki, Approximations of arbitrary relations by partial orders, Int. J. Approx. Reason. (2018), https://doi.org/10.1016/j.ijar.2018.04.012