Information Sciences 366 (2016) 48–69

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Relational division in rank-aware databases

Ondrej Vaverka*, Vilem Vychodil

Department of Computer Science, Palacky University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic

ARTICLE INFO

Article history:
Received 8 July 2015
Revised 11 January 2016
Accepted 28 February 2016
Available online 20 May 2016

Keywords:
Relational databases
Division operation
Fuzzy logic
Ranking
Imprecise queries

ABSTRACT

We present a survey of existing approaches to relational division in rank-aware databases, discuss issues of the present approaches, and outline generalizations of several types of classic division-like operations. We work in a model which generalizes the Codd model of data by considering tuples in relations annotated by ranks, indicating degrees to which tuples in relations match queries. The approach utilizes complete residuated lattices as the basic structures of degrees. We argue that unlike the classic model, relational divisions are fundamental operations which cannot in general be expressed by means of other operations. In addition, we compare the existing and proposed operations and identify those which are faithful counterparts of universally quantified queries formulated in relational calculi. We introduce Pseudo Tuple Calculus in the ranked model which is further used to show mutual definability of the various forms of divisions presented in the paper.

1. Introduction

In this paper, we present a survey and new results in the area of division-like operations in rank-aware relational models of data. In particular, we are interested in models which allow *imperfect matches of queries* in addition to the usual precise yes/no matches of queries. By an "imperfect match" we mean a situation where given record in a database does not match a query in the usual sense but the record is sufficiently close to a (hypothetical) record that matches the query exactly. In many situations, it is desirable to include records with imperfect matches in the result of a query and introduce *scores* which indicate the degrees to which the records match the given query. For instance, in a database of products, we may query for products with price equal to $1200. In the traditional understanding, a product sold for $1198 does not match the query. Nevertheless, we may want to include such product in the result and annotate it with a high score indicating that the product matches the query "almost perfectly" but not fully. In fact, reasoning with imperfect matches is inherent to human thinking and human perception of concepts like the proximity of values. Rank-aware databases [36,39] and related models of data aim at such reasoning with imperfect matches and are concerned with its formalization, analysis, and implementation in computer database systems.

Our investigation of division-like operations is motivated by the fact that in most of the existing rank-aware approaches to databases, discussion of such operations is either completely omitted or focuses only on particular Codd-style divisions. Indeed, compared to operations like projections and joins, the current rank-aware approaches pay little or no attention to division-like operations.

There seem to be two reasons for the absence of discussions of divisions in rank-aware models: First, a proposed rank-aware model simply omits divisions because its authors do not consider such an operation important. Second, the authors

* Corresponding author. Tel.: +420 585 634 715; fax: +420 585 634 716.
E-mail address: ondrej.vaverka@upol.cz (O. Vaverka).

http://dx.doi.org/10.1016/j.ins.2016.02.060
0020-0255/

of a rank-aware model expect a division-like operation to be definable by the remaining operations in a similar way as in the classic relational model of data. We argue that neither of the points is tenable and divison-like operations deserve our attention:

(1) *Divisions are important.* Division-like operations are considered in relational query systems in order to express queries which take form of particular categorical propositions. It is well understood that classic relational queries of the form "some φ is ψ" can be expressed by means of combinations of projections and natural joins which are known as semijoins. Analogous queries can also be considered in rank-aware approaches with the same meaning except for the fact that the results of queries are annotated by scores. Naturally, one should expect to be able to formulate queries of the form of categorical proposition "all φ are ψ" in a rank-aware model. In the classic model, such queries are expressed by division-like operations. In addition, some variants of the classic relational division have a close relationship to the notions of containment (subsethood) of relations. From this viewpoint, one should expect that containments and divisions in a rank-aware model should both be defined and related as in the classic model. Note that division-like operations are also interesting from the data-analytical point of view. For instance, concept-forming operators in formal concept analysis [29] can be seen as particular relational divisions.

(2) *Divisions in rank-aware models are fundamental operations.* If a rank-aware model contains operations of difference (relational minus), projection, and natural join, one may argue that a Codd-style division [16] is a definable operation in the ranked model in much the same way as it is definable in the classic model. While in the classic model, reasonable division-like operations can indeed be derived, we show further in the paper that this assumption cannot be universally adopted in rank-aware models. Technically, the operation can be defined as in the ordinary case but in many cases it lacks the basic properties of "reasonable division" and no longer is a faithful representation of queries of the form of categorical propositions "all φ are ψ". As a matter of fact, we argue in the paper that suitable variants of divisions (or equivalent formalisms) should be included as fundamental operations in rank-aware models.

In this paper we focus on divisions from the perspective of a relational model which can be seen as a generalization of the Codd [16] model of data from the point of view of residuated structures of degrees. The basic idea of the model is that tuples in relations are annotated by scores indicating degrees to which tuples match queries analogously as in [26,27], cf. also [39] introducing RankSQL and a survey paper [36]. Our model differs in how we approach the structures of scores and, consequently, the underlying logic of imperfect matches. We use structures of degrees which are recognized by fuzzy logics in the *narrow sense* [14,15,28,31,32] and the principle of *truth functionality* because our intention is to develop the model so that particular issues handled in the model (like querying and data dependencies) can be analyzed in terms of logical deduction in the narrow sense. This is in contrast with various approaches that appeared earlier [8,9,11,23] and utilized techniques from fuzzy sets (in the wide sense) where the connection to residuated structures of degrees is not so strict. We argue in the paper that the role of residuated structures is crucial for a sound treatment of division-like operations.

Our paper is organized as follows. In Section 2, we recall basic notions of our model. In Sections 3 and 4, we survey existing and propose new approaches to division operations in the classic as well as in the graded setting. In Section 5, we introduce a query language called Pseudo Tuple Calculus (PTC) that enables us to reason about the operations with ease. Finally, in Section 6, we utilize PTC to derive further observations on the mutual definability of the division operations described in the paper.

2. Relational model based on residuated structures

In this section, we present a survey of utilized notions from residuated structures of degrees and fuzzy relational systems. Furthermore, we introduce the basic notions of the generalized relational model of data and its relational algebra [4–6].

2.1. Structures of degrees

We use complete residuated lattices as structures of degrees which represent scores assigned to tuples and indicating degrees to which tuples match queries. A *residuated lattice* [3,28,32] is a general algebra [45] of the form

$$\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle \tag{1}$$

such that $\langle L, \wedge, \vee, 0, 1\rangle$ is a bounded lattice [7] with 0 and 1 being the least and the greatest element of L, respectively; $\langle L, \otimes, 1\rangle$ is a commutative monoid (i.e., $\otimes$ is commutative, associative, and $a \otimes 1 = 1 \otimes a = a$ for each $a \in L$); $\otimes$ (a multiplication) and $\rightarrow$ (a residuum) satisfy the *adjointness property*:

$$a \otimes b \leq c \text{ iff } a \leq b \rightarrow c \tag{2}$$

for each $a, b, c \in L$ where $\leq$ is the order induced by the lattice structure of $\mathbf{L}$ (i.e., $a \leq b$ iff $a = a \wedge b$). A residuated lattice (1) is called *complete* if its lattice part is a complete lattice, i.e., if L contains infima (greatest lower bounds) and suprema (least upper bounds) of arbitrary subsets of L. The multiplication $\otimes$ and its adjoint residuum $\rightarrow$ can be seen as general aggregation functions which interpret general "conjunction" and "implication" of scores, respectively. That is, if a tuple matches query Q_1 with a score a_1 and it also matches query Q_2 with a score a_2, then $a_1 \otimes a_2$ may be interpreted