Enriching the ER model based on discovered association rules

Guoqing Chen *, Ming Ren, Peng Yan, Xunhua Guo
School of Economics and Management, Tsinghua University, Beijing 100084, China
Received 2 July 2006; accepted 3 July 2006

Abstract

The entity–relationship (ER) model, a powerful means for business and data modeling, needs to be enriched with new semantics as the real world changes and its understanding improves. This paper attempts at enriching the ER model based on association rules (AR) discovered from large databases by introducing specializations and sub-types into the ER model. The proposed framework is extended to deal with more general, flexible and linguistic knowledge in fuzzy association rules. Moreover, transforming an AR-enriched-ER (AR-EER) schema to a relational database (RDB) schema is also investigated.
© 2006 Elsevier Inc. All rights reserved.

Keywords: The entity–relationship model; Association rule; Specialization; Fuzzy sets

1. Introduction

As applications of computer-aided design and manufacturing (CAD/CAM) become increasingly widespread in industrial engineering, an efficient management of large amounts of product, operation and process data in databases becomes crucial for improving design and production quality as well as for effective decision making. The entity–relationship (ER) model [7] has evolved considerably since its inception, and been widely accepted as a standard tool for conceptual modeling of relational databases. Basic ER concepts, namely entity, relationship and attribute, have been extended in several directions, resulting in a number of extensions such as aggregation, refinement, specialization/generalization, inheritance, etc. [9,10,15]. Fuzzy ER models [4,13,19] have been proposed to deal with different types of uncertainty, which largely improves expressiveness and usefulness of ER. An ER schema, built upon the knowledge available from business managers and systems analysts, needs to keep extended or enriched with new semantics as the real world concerned changes and/or the understanding of the real world improves. In fact, the idea that hidden relationships should be considered in ER modeling has been mentioned in [8]. This paper focuses on enriching an ER schema based on semantics..
obtained from data mining, which is implicit, previously unknown, and potentially useful knowledge in large databases.

A particular type of knowledge of concern from data mining is association rule (AR) [1], which reflects an association among a collection of items \( I = \{I_1, I_2, \ldots, I_m\} \). An AR with two items comes in the form \( I_i \Rightarrow I_j, i, j \in \{1, 2, \ldots, m\}, i \neq j \), which means that “occurrence of \( I_i \) is associated with occurrence of \( I_j \)”. Its two basic measures, degree of support (Dspp(I \(_i\) \( \Rightarrow \) I \(_j\)) = \(|\{I_i, I_j\}| / |D|\)) and degree of confidence (Dconf(I \(_i\) \( \Rightarrow \) I \(_j\)) = \(|\{I_i, I_j\}| / |\{I_i\}|\)), should be equal to or greater than pre-specified thresholds (called min-support and min-confidence), where |\( D \)| is the number of transactions in database \( D \), |\{I\(_i\)\}| is the number of transactions that contain item \( I_i \) and |\{I, I\(_j\)\}| is the number of transactions that contain items \( I_i \) and \( I_j \). Item \( I_k \) can be represented in the form of a tuple \( \langle P, p \rangle \), where \( P \) is a descriptive property of an object, and \( p \) is a singleton or a subset of the domain of \( P \). For example, a rule \( \langle \text{category}, \text{furniture} \rangle \Rightarrow \langle \text{age}, [35, 55] \rangle \) reflects an association between category of products and age of buyers. Various forms of association rules have been developed such as generalized, multi-level, and quantitative association rules [12, 16, 17]. Fuzzy association rules [2, 5, 6] have also been proposed to facilitate linguistic expressions. As far as the ER modeling is concerned, the semantics reflected by an AR may either be unaware of at the time when the ER schema was built (e.g., due to the limitation of the scope of expertise), or be previously unimportant/uninteresting but now become significantly meaningful. It is worth mentioning that an AR needs to be inspected by domain experts before being considered in the ER enrichment, which however goes beyond the scope of this paper.

In this paper, we present a framework that bridges the paradigms of ER and AR. Here AR is used to enrich an ER schema by introducing a new construct called relationship specialization (i.e., R-specialization), giving rise to the AR-enriched-ER (AR-EER) model. Furthermore, an AR-EER schema as an enriched conceptual model will be transformed with an extended Crep algorithm into a relational database schema. It will be proven that the relational database schema has the equivalent information capacity to that of the AR-EER schema, and is in the normal form of BCNF. Moreover, since fuzziness is a type of uncertainty inherent in human reasoning and decision-making processes, the framework will be discussed in the context of fuzzy ER and fuzzy AR. The paper is organized as follows. Section 2 introduces the framework of enriching an ER schema based on discovered ARs. In Section 3, an AR-EER schema is formulated and transformed into a relational database schema, which is proven to preserve the information capacity of the ER schema, and is in BCNF. Section 4 discusses fuzziness in data models and discovered knowledge to enrich semantic representation of the AR-EER model.

2. The AR-enriched ER (AR-EER) model

2.1. E-specialization and R-specification

The ER model describes (static) aspects of the real world by structuring entity types and relationship types between them. An entity/relationship type, defined as a set whose instances are entities/relationships, is characterized by a set of attributes \( A(\cdot) \), with each attribute associated with a value set \( \text{Dom}(\cdot) \). An attribute-based specialization (interchangeably called specialization in the paper) is traditionally referred to as a process of defining sub-types on an entity type (i.e., \( E \)-specialization), in which instances take the same value on the given attribute, or take values from the same partition of the value sets. For example, in Fig. 1(a), three sub-types are defined on attribute \( A_1 \) of entity type \( E_1 \) (i.e., \( A_1 \in A(E_1) \)) with three specific values \( v_1, v_2, \) and \( v_3 \), respectively. For the sake of simplicity, such \( v_k \) values (e.g., \( k = 1, 2, 3 \) in Fig. 1(a)) are supposed to be subsets of the attribute domain concerned in the following discussion. \( E_1 \) is called the super-type of \( E_{1k}, E_{1k} \subseteq E_1 \). Formally we have \( E_{1k} = \{ e | V_{A_1}(e) \in v_k, e \in E_1 \} \) where \( V_{A_1}(e) \) is the value entity \( e \) takes on attribute \( A_1 \). Furthermore, relationships may be sub-grouped as entities can be, by taking specific values on the given attribute, which gives rise to the notion of \( R \)-specialization. Fig. 1(b) shows a \( R \)-specialization, where \( R_0 \) is a sub-type defined on attribute \( A_0 \) (\( A_0 \in A(R) \)) with specific values in \( v_0 \). The dotted lines indicate its participating entity types.

\(^1\) Relational normal forms are properties defined for relational database schemes, aimed at reducing data redundancy and avoiding update anomaly. Boyce–Codd Normal Form (BCNF) requires all functional dependencies to be key dependencies.
Formally, we have $R_0 = \{ r | V_{A_0}(r) \in v_0, \, r \in R \}$ and $R_0 \subseteq R$. R-specialization can be defined not only on an attribute of a relationship type, but also on an attribute of the participating entity type. In Fig. 1(c) $R_0$ is defined on attribute $A_1$ of $E_1$, as a sub-grouping of instances of $R$ with participants in $E_{11}$ and $E_2$. Formally we have $R_0 = \{ r = (e_1, e_2) | V_{A_1}(e_1) \in v_0, \, e_1 \in E_1, \, r \in R \}$ and $R_0 \subseteq R$.

As an example, suppose $R$ is a relationship type Order linking entity types Product ($E_1$) and Buyer ($E_2$), Order and Product are associated with attributes amount ($P$) and price ($A$), respectively. A relational database can be designed based on the schema, and populated then-after whenever transactions take place. Suppose association rules have been discovered from the databases and the following three properties of an AR item are regarded as new knowledge with semantics suited to enrich the ER schema.

2.2. Enriching the ER schema based on association rules

A property of an AR item is derived from a relational attribute in a database, which can be mapped to an attribute of some type in the ER schema. Concretely, $\langle P, p \rangle$, $\exists E$ and $A_1 \in A(E)$ (or $\exists R$ and $A_j \in A(R)$) such that $P = E.A_1$ (or $P = R.A_j$), and $p \subseteq \text{Dom}(A_i)$. In this way, an AR item characterizes a group of instances of a type which take specific values on an attribute $A_i$ and leads to a specialization on attribute $A_i$. For example, $\langle \text{price}, [10k, 15k) \rangle$ may lead to a specialization of Product by sub-grouping products with prices between 10k and 15k. Generally, for an AR $\langle P_1, p_1 \rangle \Rightarrow \langle P_2, p_2 \rangle$, three cases of R-specialization can be categorized.

An ER schema of logistics in a supply chain (Fig. 3) is taken as an illustrative example. Two types of employees, namely bargainer and salesman, deal with contracts with vendors and sales to retailers, respectively. A relational database can be designed based on the schema, and populated then-after whenever transactions take place. Suppose association rules have been discovered from the databases and the following three are regarded as new knowledge with semantics suited to enrich the ER schema.
Rule 1: (Retailer.district, HD) ⇒ (Product.price, [10k, 15k]);
Rule 2: (contract.amount, [50k, 100k]) ⇒ (Vendor.corp_term, [2, 3]);
Rule 3: (supply.term, [6, 12]) ⇒ (supply.quantity, [50, 100]).

Case 1. Items relating to two entity types. Suppose \( P_1 \) and \( P_2 \) are \( E_1.A_1 \) and \( E_2.A_2 \), respectively, a \( R \)-specialization can be defined on \( R \) resulting in a relationship sub-type \( R_1 = \{ r = (e_1, e_2) | V_{A_1}(e_1) \in p_1, V_{A_2}(e_2) \in p_2, e_1 \in E_1, e_2 \in E_2, r \in R \} \), with participants \( E_{11} = \{ e | V_{A_1}(e) \in p_1, e \in E_1 \} \) and \( E_{21} = \{ e | V_{A_2}(e) \in p_2, e \in E_2 \} \). For example, rule 1 will result in a relationship sub-type HD_price [10k, 15k] linking Retailer_in_district_HD and Product_with_price[10k,15k], as shown in Fig. 4.

Case 2. Items relating to a relationship type and an entity type. Suppose \( P_1 \) and \( P_2 \) are \( R.A_0 \) and \( E_1.A_1 \), respectively, a \( R \)-specialization can be defined on \( R \) resulting in a relationship sub-type \( R_2 = \{ r = (e_1, e_2) | V_{A_0}(r) \in p_1, V_{A_1}(e_1) \in p_2, e_1 \in E_1, r \in R \} \), with participants \( E_{11} = \{ e | V_{A_1}(e) \in p_2, e \in E_1 \} \) and \( E_2 \). An example can be the relationship sub-type Amount[50k, 100k]_corp [2, 3) years in Fig. 4.
3. Fuzziness in data models and knowledge representation

Treatment of uncertainty and imprecision is one of the key issues in data modeling and knowledge representation. Since fuzzy terms and linguistic words are often involved in describing a property of AR and an attribute of ER, the above framework is extended to enable enriching the ER model based on fuzzy AR, specially dealing with the fuzzy sets on domains of attributes (e.g., high on Dom(price)).

Fuzzy ER concepts have been introduced in [4], where super-types and sub-types can be fuzzy sets. In our proposed model, original ER schema is assumed to be non-fuzzy (i.e., super-types are non-fuzzy). In a specialization, \( v_k \) can be a fuzzy set defined on the domain of an attribute, so that the corresponding sub-type is a fuzzy set, to which each instance belongs with a membership degree in [0,1]. For example, a product may belong to Expensive-product with a degree of 0.8. Look at Fig. 1(a) again in a fuzzy perspective: an entity \( e \) belongs to \( E_1 \) with a degree \( \mu_{E_1}(e) \) calculated by closeness measure [4] between \( V_{A_1}(e) \) and fuzzy set \( v_1 \). Formally, \( \mu_{E_1}(e) = \approx (V_{A_1}(e), v_1) \), and \( E_{11} = \{ (e, \mu_{E_1}(e)) | \mu_{E_1}(e) = \approx (V_{A_1}(e), v_1), e \in E_1 \} \), where \( \approx \) is a closeness measure. Likewise for Fig. 1(b), a relationship \( r \) belongs to \( R_0 \) with the degree \( \mu_{R_0}(r) = \approx (V_{A_0}(r), v_0) \), and \( R_0 = \{ (r, \mu_{R_0}(r)) | \mu_{R_0}(r) = \approx (V_{A_0}(r), v_0), r \in R \} \); in Fig. 1(c) a relationship \( r \) belongs to \( R_0 \) with the degree \( \mu_{R_0}(r) = \mu_{E_1}(e) \), and \( R_0 = \{ (r, \mu_{R_0}(r)) | \mu_{R_0}(r) = \approx (V_{A_1}(e_1), v_1), e_1 \in E_1, r = (e_1,e_2) \in R \} \). In a R-specialization with two attributes, \( r \) belongs to \( R_0 \) with a degree \( \mu_{R_0}(r) \), which may be calculated by, say, minimum of related closeness measures. For \( R_1 \), we have \( \mu_{R_1}(r) = \min(\mu_{E_1}(e_1), \mu_{E_1}(e_2)) \). Fuzzy relationship sub-types resulted from such specializations are extended as follows:

\[
R_1 = \{ (r, \mu_{R_1}(r)) | \mu_{R_1}(r) = \min(\approx (V_{A_1}(e_1), p_1), \approx (V_{A_1}(e_2), p_2)), e_1 \in E_1, e_2 \in E_2, r = (e_1,e_2) \in R \}
\]

\[
R_2 = \{ (r, \mu_{R_2}(r)) | \mu_{R_2}(r) = \min(\approx (V_{A_2}(e_1), p_1), \approx (V_{A_2}(e_2), p_2)), e_1 \in E_1, e_2 \in E_2, r = (e_1,e_2) \in R \}
\]

\[
R_3 = \{ (r, \mu_{R_3}(r)) | \mu_{R_3}(r) = \min(\approx (V_{A_3}(e_1), p_1), \approx (V_{A_3}(e_2), p_2)), r \in R \}
\]

where \( p_1, p_2 \) are fuzzy sets defined on attribute domains.

In a fuzzy association rule (FAR), \( p \) in item \( \langle P, p \rangle \) is a fuzzy set defined on Dom(P), and \( \langle P, p \rangle \) is called a fuzzy item. A fuzzy AR, which contains at least one fuzzy item, can be used to enrich an ER schema by introducing fuzzy specializations and fuzzy sub-types into an ER schema. Take a fuzzy association rule \( \langle \text{Retailer.district, HD} \rangle \Rightarrow \langle \text{Product.price, high} \rangle \) as an example, fuzzy item \( \langle \text{Product.price, high} \rangle \) will result in a fuzzy entity sub-type Expensive-product, and thus lead to a fuzzy relationship sub-type HD_expensive. An instance \( r \) belongs to HD_expensive with a degree \( \mu_{\text{HD\_expensive}}(r) = \mu_{\text{Expensive\_product}}(e_1), r = (e_1,e_2) \).

4. Relational database modeling based on the AR-EER model

A common approach for the relational database (RDB) design is to describe the constructs and constraints of applications in an ER schema, and transform the ER schema into a relational database schema, which has attracted extensive research attention [11,14,18]. Since new constructs and constraints emerge in an AR-Enriched ER schema, the transformation to a relational database schema needs to be investigated, focusing on relationship sub-types.

4.1. Transforming the AR-EER model to the relational database model

A relational database schema includes a set of relation schemes like \( R(X) \), where \( R \) is the scheme name and \( X \) is the attribute set. A relation corresponding to \( R(X) \) contains a number of tuples, and a group of relations constitute a state of a relational database schema. Such a state is consistent if it satisfies pre-defined dependencies and constraints. In [14] ER constructs as entity/relationship type and entity sub-type are first transformed.
by algorithm Crep canonically into relation schemes with constraints as: (1) functional dependency on \( R(X) \): 
\[ X_1 \rightarrow X_2, X_1, X_2 \subseteq X \]; (2) inclusion dependency on \( R(X) \) (with respect to \( R(X) \)): 
\[ R_1(X_0) \subseteq R(X_0) \] and (3) nulls-not-allowed constraint on \( R(X) \): \( \emptyset \rightarrow X \), where attributes in \( X \) are not allowed to have null values. The relation schemes are then normalized by algorithm Norm into relation schemes in BCNF. The schema transformation of the AR-EER schema follows three steps.

Step 1. Formulating the consistent state of an AR-EER schema.

A type is associated with a set of local attributes \( A(\cdot) \), which are supposed to have atomic and non-null values, and universally unique names. Consider sub-type \( R_1 \) as an example. Types \( E_1, R, E_{11} \) are represented by \( A(E_1), A(R) \cup A^F(R) \), and \( A(E_{11}) \cup A^F(E_{11}) \), respectively, where \( A^F(R) = A(E_1) \cup A(E_2), A^F(E_{11}) = A(E_1) \); \( R_1 \) is represented by \( A(R_1) \cup A^F(R_1) \), where \( A^F(R_1) = (A(R) \cup A^F(R)) \cup A(E_{11}) \cup A(E_{21}) \). For example, if 
\[ A(\text{Order}) = \{K_0, A_0\}, A(\text{Product}) = \{K_1, A_1\}, A(\text{Retailer}) = \{K_2, A_2\}, A(\text{Product_with_price}[10k, 15k]) = \{V_1\}, \]
\[ A(\text{Retailer_in_district}_{\text{HD}}) = \{V_2\}, A(\text{HD_price}[10, 15k]) = \{V_0\}, \]
then \( A^F(\text{HD_price}[10, 15k]) = \{K_0, A_0, K_1, \]
\[ A_1, K_2, A_2, V_1, V_2\}. \]

At the semantic level, a collection of instances constitutes a state of the ER schema. A state is consistent if the instances satisfy the constraints as \([14]\): (1) Identification: \( E_1 \) and \( R \) are identified by \( A^F(E_1) \) and \( A^F(R) \cup A^F(R) \), respectively, \( A(\cdot) \subseteq A^F(\cdot) \); sub-types are identified by \( A^F(\cdot) \); and (2) Cardinality: If the cardinality between \( E_1 \) and \( E_2 \) is 1:1 or 1:M, then each entity in \( E_1 \) corresponds to a distinct entity in \( E_2 \).

Step 2. Transforming relationship sub-types canonically into relation schemes with constraints.

The proposed method, namely extended Crep (i.e., ECrep), deals with entity/relationship type and entity sub-type as the Crep method does, and also transfers a relationship sub-type into a relation scheme as follows: \( R_1 \) is transformed into \( R_1(X) \), where relational attributes in \( X \) are in an one-to-one correspondence with ER attributes in \( A(R_1) \cup A^F(R_1) \), and the domain of a relational attribute corresponds to the value set of the corresponding ER attribute. Attributes in \( X_1 \) correspond to \( A^F(R_1) \), with subsets \( X_{1k} \ (k = 1, 2, 3) \) corresponding to \( A(R) \cup A^F(R), A(E_{11}) \cup A^F(E_{11}) \) and \( A(E_{21}) \cup A^F(E_{21}) \), respectively. The constraints on \( R_1(X) \) are:

\[
\text{FD-1: } X_1 \rightarrow (X - X_1); \quad \text{IND-1: } R_1(X_{11}) \subseteq R(X), \quad R_1(X_{12}) \subseteq R_{E_1}(X), \quad R_1(X_{13}) \subseteq R_{E_2}(X); \quad \text{NN-1: } \emptyset \rightarrow X
\]

where \( R(X), R_{E_1}(X) \) and \( R_{E_2}(X) \) are relation schemes transformed from \( R, E_{11} \) and \( E_{21} \), respectively.

At the semantic level, relationships of \( R_1 \) are mapped to tuples of \( R_1(X) \) and reversely.

\( \rho \): Each relationship \( r \) of \( R_1 \) is mapped into a tuple \( t \) of \( R_1(X) \), such that ER attribute value \( V_{A_0}(r) \), 
\( A_0 \in A(R_1) \cup A^F(R_1) \) is mapped into relational attribute value \( t(A_0), A_0 \in X \);

\( \rho' \): Each tuple \( t \) of \( R_1(X) \) is mapped into a relationship \( r \) of \( R_1 \), such that relational attribute value \( t(A_0) \), 
\( A_0 \in X \), is mapped into an ER attribute value \( V_{A_0}(r) \), \( A_0 \in A(R_1) \cup A^F(R_1) \).

Suppose Order, Product, Retailer, Product_with_price[10k, 15k], and Retailer_in_district_HD are transformed into relation schemes \( R_{O}(K_0A_0K_1A_1K_2A_2), R_{P}(K_1A_1), R_{RT}(K_2A_2) \), \( R_{PO}(K_1A_1V_1) \) and \( R_{RT-1}(K_2A_2V_2) \), respectively, then we have the relationship sub-type HD_price[10k, 15k] transferred into a relation scheme \( R_{O-1}(K_0A_0K_1A_1K_2A_2V_1V_2V_0) \) with constraints:

\[
\text{FD-1: } K_0A_0K_1A_1K_2A_2V_1V_2 \rightarrow V_0 \quad \text{IND-1: } \begin{array}{l}
R_{O-1}(K_0A_0K_1A_1K_2A_2) \subseteq R_{O}(K_0A_0K_1A_1K_2A_2) \quad \emptyset \rightarrow K_0A_0K_1A_1K_2A_2V_1V_2V_0 \\
R_{O-1}(K_1A_1V_1) \subseteq R_{PO}(K_1A_1V_1) \\
R_{O-1}(K_2A_2V_2) \subseteq R_{RT-1}(K_2A_2V_2)
\end{array}
\]

\( ^2 \) According to Crep, FD defined on \( R_{O}(K_0A_0), R_{P}(K_1A_1), R_{RT}(K_2A_2) \) are \( K_i \rightarrow A_i \), \( i = 1, 2, 3 \), respectively.
Step 3. Normalizing the relation schemes to get a schema in the normal form of BCNF.

It should be indicated that \( FD-1 \) is a partial dependency due to a set of redundant attributes in \( X_1 \), for example, \( A_0A_1A_2V_1V_2 \) in attribute set of \( R_{O-1} \), which can be removed by \( Norm \). Apply \( Norm \) to \( R_{O-1} \), the relation scheme is updated to \( R_{O-1}(K_0K_1K_2V_0) \) with dependencies and constraints as follows:

\[
 FD-1': K_0K_1K_2 \rightarrow V_0 \quad IND-1': R_{O-1}(K_0K_1K_2) \subseteq R_{O}(K_0K_1K_2) \quad NN-1': \emptyset \rightarrow K_0K_1K_2V_0
\]

**Theorem 1.** Let an AR-EER schema \( S \) be transformed into a relational schema and then normalized by \( Norm \) into a relational schema \( RS \). Then \( RS \) has the equivalent information capacity to that of \( S \).

**Proof.** The proof focuses on transformation of relationship sub-types in terms of conditions of equivalent information capacity, since \( Norm \) and \( Crep \) have been proven to be information-capacity preserving. Let \( S_1 \) denote the set of relationship sub-types in \( S \) and be transformed into a set of relation schemes \( S_2 \).

(a) Given a consistent state of \( S_1 \), the resultant state of \( S_2 \) is proved consistent inductively. Suppose the initial state of \( S_2 \) is empty, and the inductive hypothesis is that before a relationship \( r \) of relationship sub-type \( R_1 \) in \( S_1 \) is mapped by \( \rho \), the state of \( S_2 \) is consistent. We apply \( \rho \) to \( r \), and we get a tuple \( t \). Since \( r \) is known to satisfy the dependencies and constraints of \( S \), \( t \) satisfies \( FD-1 \), \( IND-1 \) and \( NN-1 \) and \( S_2 \) satisfies the dependencies and constraints after \( r \) is mapped. Apply \( \rho \) to other relationships until relationships in \( S_1 \) are all mapped into tuples in \( S_2 \).

(b) The inductive proof applies by analogy and is omitted due to the limitation of space.

(c) \( \rho \) maps \( r \) to \( t \), with \( r \) and \( t \) in a one-to-one correspondence of identical values, so \( \rho \) maps distinct relationships to distinct tuples; by analogy \( \rho' \) maps distinct tuples to distinct relationships. So \( \rho \cdot \rho' \) is the identity on consistent states of \( S_1 \), and \( \rho' \cdot \rho \) is the identity on consistent states of \( S_2 \).

(d) Condition (d) is satisfied according to the definition of \( \rho \) and \( \rho' \).

**Theorem 2.** Relation schemes transformed by \( ECrep \) and normalized by \( Norm \) are in normal form of BCNF.

**Proof.** Suppose a relationship sub-type \( R_1 \) is transformed and normalized into \( R_1(\bar{X}) \) with \( FD-1': K \rightarrow (X - K) \), which is a nontrivial functional dependency. According to the definition of \( Norm \), \( K \) as the super-key of \( R_1(\bar{X}) \) is the super-key of \( R_1(X) \), while any sub-set of \( K \) is not. Hence \( FD-1' \) is a key dependency. The relation scheme \( R_1(X) \) is proved to be in BCNF.

4.2. Transforming the fuzzy AR-EER model to a fuzzy-relation-based framework

Researchers have proposed different fuzzy relational database models, among which fuzzy-relation-based framework [3] generalizes the typical relational database model by accepting degrees of tuple-belongings in \([0,1]\) instead of only in \(\{0,1\} \). Formally, the general form of a fuzzy relation scheme \( R \) can be express as \( R = \{(t_1, \mu_R(t_1)), (t_2, \mu_R(t_2)), \ldots, (t_n, \mu_R(t_n))\}, \mu_R(t_i) \in [0,1] \), and the degree of tuple belonging \( \mu_R(t_i) \) is the membership degree that \( t_i \) belongs to \( R \). At the logical level, the fuzzy-relation-based framework is employed to handle the fuzziness.

The \( ECrep \) method may be further extended for the fuzzy case to deal with the degree \( \mu_{R_1}(r) \) (and \( \mu_{E_1}(e), \mu_{E_2}(e) \)) as an attribute of \( R_1 \) with a value set \([0,1]\). Then fuzzy sub-type \( R_1 \) is transformed to a fuzzy relation scheme \( R_1(X) \) with a “relational attribute” as \( \mu_{R_1}(t) \), with the attribute domain \([0,1]\).

---

3 The two schema \( S_1, S_2 \) are said to have equivalent information capacity if there exist total functions \( \rho \) and \( \rho' \) such that [14]: (a) \( \rho \) maps consistent states of \( S_1 \) into consistent states of \( S_2 \); (b) \( \rho' \) maps consistent states of \( S_1 \) into consistent states of \( S_2 \); (c) the composition of \( \rho \) followed by \( \rho' \) is the identity on the set of all consistent states of \( S_1 \); the composition of \( \rho' \) followed by \( \rho \) is the identity on the set of all consistent states of \( S_2 \) and (d) \( \rho \) preserves the data values of any state of \( S_1 \), and \( \rho' \) preserves the data values of any state of \( S_2 \).
At the semantic level, relationships of $R_1$ are mapped to tuples of $R_1(X)$ as in the crisp case, with their membership degrees mapped to the degree of the tuple belonging as follows:

\[ \rho: \text{Each relationship of } R_1 \text{ is mapped into a tuple of } R_1(X), \text{ and the degree } \mu_{R_1}(r) \text{ is mapped into a tuple belonging } \mu_{R_1}(t); \]

\[ \rho': \text{Each tuple of } R_1(X) \text{ is mapped into a relationship of } R_1, \text{ and the tuple belonging } \mu_{R_1}(t) \text{ is mapped into } \mu_{R_1}(r). \]

For example, HD\_expensive in Section 3 will result in a fuzzy relation scheme $R_{O-2} = \{(t_1, \mu_{O-2}(t_1)), \ldots, (t_n, \mu_{O-2}(t_n))\}$, where $t_i (i = 1, 2, \ldots, n)$ is a collection of values of $K_0K_1K_2V_0$ and $\mu_{O-2}(\cdot)$ is tuple belonging of $t_i$.

5. Conclusions

The paper has proposed an enriched ER model based on discovered association rule (AR) to better reflect semantics of the domain concerned, by introducing specializations and sub-types into an ER schema, resulting in an AR-Enriched ER (AR-EER) schema. The framework has also been extended to facilitate representation of imprecise information and partial knowledge in the ER model in accordance with linguistic expressions and fuzzy association rules. Moreover, an AR-EER schema has been transformed by an extended Crep algorithm into a relational schema, which has proven to have equivalent information capacity to that of the AR-EER schema, and is in BCNF.

References
