AN EXTENSION OF FUZZY RELATIONAL DATABASE MODEL INTO VAGUE RELATIONAL DATABASE MODEL

Synopsis submitted to

WEST BENGAL UNIVERSITY OF TECHNOLOGY

In partial fulfillment of the Degree of

DOCTORATE OF PHILOSOPHY

2013

Jaydev Mishra

SCHOOL OF ENGINEERING AND TECHNOLOGY
WEST BENGAL UNIVERSITY OF TECHNOLOGY
SALTLAKE, BF-142, SECTOR-1, KOLKATA-700 064
WEST BENGAL
INDIA

Contents

1. Introduction and Motivation

Relational database systems have been extensively studied worldwide since Codd [3] had proposed the relational data model in 1970. Based on this model, several commercial relational database systems are also available. This data model usually takes care of precisely defined and unambiguous data. However, in the real world applications data are often partially known i.e., incomplete or imprecise. For example, instead of specifying that the height of David is 188 cm, one may say that the height of David is around 190 cm, or simply that David is tall. Other examples on uncertain data may be "Salary of Ram is around Rs. 60000", "Ram has a high salary", "Salaries of almost equally experienced employees are more or less same" etc. All these are informative statements that may be useful in answering queries or making inferences. However, such type of data cannot be represented in the classical relational data model. In order to incorporate such imprecise or fuzzy data, the classical relational data model has been extended by several authors on the mathematical framework of **fuzzy set theory** which was initially introduced by Zadeh [11] in 1965. The idea of fuzzy set theory is to assign a single membership value $\mu_F(u) \in [0,1]$ to any element $u \in U$ of a fuzzy set F where U is universe of discourse. $\mu_F(u)$ is called the grade of membership of the element u in the set F. Such an extended database model is called a fuzzy relational database model which can represent and manipulate uncertain or vague information.

The study of different integrity constraints and the associated inference rules constitute a major part of research in any database design. Different type of integrity constraints such as functional dependency, multivalued dependency, join dependency etc. have been identified and studied widely in literature which are considered to be crucial for the purpose of designing a database. Another primary objective of any good database design is to decrease data redundancy and to provide data reliability. Such data redundancies and insertion, deletion and updation of anomalies have been of great concern in a relational database. Thus, like integrity constraints, normalization process for minimizing redundancy and data anomalies also plays an important role in designing a good relational database. Consequently, as one extends the classical relational database model into a fuzzy relational data model, it becomes necessary to study the integrity constraints in the light of fuzzy concept and several authors [1, 2, 7-10] have contributed in this direction. It is also essential to design normalized database schemes to minimize data redundancies and different anomalies for fuzzy relational models [2, 8].

Vague set theory has been subsequently introduced by Gau and Buehrer [4] in 1993 as a more efficient tool to treat ambiguous data and it has been successfully applied in many different fields such as fuzzy control, decision analysis, expert system etc. A vague set, conceived as a generalization of the concept of fuzzy set, is a set of decision objects, each of which has a grade of membership whose value is a continuous subinterval of [0,1]. It is characterized by a truth-membership function and a false-membership function and separates the positive and negative evidence for membership of an element in the set. Thus a vague set has more powerful ability to process imprecise information than traditional fuzzy sets. Therefore classical relational databases may also be extended to represent and deal with uncertain data in a more effective manner with the idea of vague set theory. Such an extended database model is then called a vague relational database model. Research in this direction has attracted some attention in recent years. However, till date very limited number of references [5, 6, 12-14] are available in this area. In particular, a complete theory for the design of a vague relational database has not yet been reported in literature. Hence the present work has been primarily aimed at development of the design theory of a vague relational database model.

2. Objectives and Scope of Present Research

The main objective of the present work is to design a vague relational database model that can treat uncertain information and queries and compare its utility and effectiveness with the more conventional fuzzy database model. Thus the work reported in this thesis begins with the design a fuzzy relational database model that can accommodate real world data and imprecise queries and then all the relevant theories have been extended with the concept of vague set theory. The design theory of fuzzy data model is mainly based on a definition of fuzzy functional dependency introduced by Al-Hamouz and Biswas [1] on the basis of equivalence relation. This definition of **ffd** (called α -**ffd**) which uses the concepts of α -nearer elements and α **-**equality of tuples made the development of the design theory of fuzzy database easier compared to the definition of **ffd** given by other authors. In the existing concepts of **ffd,** the comparison of two data in a domain is done with the help of fuzzy equality relations, which are not equivalence relations. This notion of α -**ffd** has been used in the present work to define partial fuzzy functional dependency (partial **ffd**), fuzzy key, fuzzy closure of an attribute set, redundant fuzzy functional dependency (redundant **ffd**). Next, we have proposed a **new definition** of fuzzy multivalued dependency (**fmvd**), called α -**fmvd** on the basis of the idea of α -equality of tuples as defined in [1]. This definition provides a very easy and straightforward way of extending **mvd** to **fmvd** for a fuzzy relational database model and thus it differs from the existing definitions in literature. A set of sound and complete inference axioms have also been proposed and proved for our α-**fmvd**. In order to design a fuzzy relational database model by minimizing data redundancies and different anomalies, we have introduced a number of fuzzy normal forms such as fuzzy first normal form (**F1NF**), fuzzy second normal form (**F2NF**), fuzzy third normal form (**F3NF**) and fuzzy Boyce Codd normal form (**FBCNF**) based on our α-**ffd** and all these concepts have been verified with suitable real life examples. Since the process of normalization based on **ffd**s uses the idea of decomposition to achieve the desired normal form, so one should also confirm dependency preservation and lossless join properties of decomposition for designing a good database model. Hence the algorithms for dependency preservation and lossless join decomposition into fuzzy third normal form (**F3NF**) have also been explained herewith**.**

In the next part of the thesis, we have made an attempt to extend this fuzzy relational database model into a vague relational database model since vague sets have an extra edge over fuzzy sets. Similar to the theory of classical and fuzzy relational databases, vague functional dependency and vague multivalued dependency may also be used as a guideline for the design of a vague relational schema. In the present work, we have introduced **new definitions** of vague functional dependency (called α-**vfd**) as well as vague multivalued dependency (called α-**vmvd**), on the basis of α-equality of tuples and the notion of similarity measure between vague sets. A set of complete and sound vague inference rules similar to the classical case have also been designed and verified for the α-**vfd** and the α-**vmvd**. **Partial** α-**vfd** and **vague key** have also been studied with the new notion of α-**vfd** and tested with examples. Further, these concepts have been used to define different vague normal forms such as vague first normal form (**V1NF**), vague second normal form (**V2NF**), vague third normal form (**V3NF**) and vague Boyce Codd normal form (**VBCNF**) in a vague relational database. Again normalization process based on **vfd**s uses a number of decompositions while normalizing the relations. But normal forms do not always guarantee a good database design. The normalization process should also confirm the existence of two additional and desirable properties, dependency preservation and lossless join property. Hence the algorithms for dependency preservation and lossless join decomposition into vague third normal form (**V3NF**) have also been presented in this part.

In the final or third part of the thesis, we have proposed an algorithm that defines an attribute independent membership function of a fuzzy set for calculating membership value of different fuzzy or vague attributes. Membership functions used previously in literature are found to be dependent on the particular attribute. Also we have designed an architecture to process uncertain queries for both fuzzy set and vague set and have observed that vague sets give more accurate result in comparison to fuzzy sets.

3. Proposed Contents of the Thesis

Chapter 1: General Introduction
Chapter 2: Fuzzy Functional Dependency
Chapter 3: Fuzzy Multivalued Dependency
Chapter 4: Normalization in Fuzzy Relational Database
Chapter 5: Vague Functional Dependency
Chapter 6: Vague Multivalued Dependency
Chapter 7: Normalization in Vague Relational Database
Chapter 8: Uncertain Query Processing
Chapter 9: Concluding Remarks

4. Brief Description of the Present Work

The research work reported in this thesis is devoted to the design of a fuzzy relational database model and its extension to a vague database model with the introduction of vague set theory. As shown in the previous section, the work has been organized in nine Chapters. Here we describe briefly the work presented in different chapters.

Chapter 1 is an introductory chapter in which the motivation and scope of the present work have been stated. A brief description of relevant mathematical theory, namely, fuzzy set and vague set as well the classical relational database model is also given. Further, a brief review of research work reported in literature in the areas of fuzzy relational database and vague relational database has also been included in the same chapter.

While extending the design theory of relational database system to the fuzzy domain, it is necessary to study fuzzy data dependencies and the associated implication problems since data dependency plays a vital role in any logical database design. Functional dependency of one set of attributes upon another is one of the most important concepts among different data dependencies in relational database. In **Chapter 2,** we have concentrated on the extension of functional dependencies in fuzzy paradigm. The present work is based on a notion of **ffd** introduced by Al-Hamouz and Biswas [1] in 2006. This definition of **ffd** uses the concept of α-equality of tuples which makes the development of the theory of fuzzy database easier compared to the definitions of **ffd** available in literature. In this chapter**,** the **f-Transitive**, **f-Union**, **f-Pseudotransitive** and **f-Decomposition** rules of [1] have been modified with proper verification. We have also extended the work to define partial fuzzy functional dependency (partial **ffd**), fuzzy key and redundant fuzzy functional dependency (redundant **ffd**) in the same chapter.

In **Chapter 3**, we present a **new** definition of fuzzy multivalued dependency (**fmvd**), called α-**fmvd,** on the basis of the α**-**equality of tuples as defined in [1]. The new definition provides a very easy and straightforward way of extending **mvd** to **fmvd** for a fuzzy relational database model and thus it differs from the existing definitions in literature. Next the definition is shown to be consistent i.e., it reduces to that of classical multivalued dependency (**mvd**) when the choice parameter α takes the value one. Then a set of sound and complete inference axioms have been proposed and proved for the α-**fmvd**.

Chapter 4 deals with fuzzy extensions of different normal forms for our fuzzy relational database model. Firstly, we have designed an algorithm to find the fuzzy closure of an attribute set which can be utilized to find fuzzy key. Next, with the concepts of α-**ffd** and partial α-**ffd** as discussed in chapter 2, we have defined various fuzzy normal forms namely, fuzzy first (**F1NF**), fuzzy second (**F2NF**), fuzzy third (**F3NF**) and fuzzy Boyce Codd (**FBCNF**) normal forms. Fuzzy prime and non prime attributes have also been defined in the same chapter. Algorithms for dependency preservation and lossless join properties for the decomposition of a fuzzy relation into the desired fuzzy normal form have also been presented in this chapter. Finally, in the same chapter we have also included a real life application which depicts how normalization based on α-**ffd**s of fuzzy relation can be done.

Since vague sets have been introduced to deal with imprecise information in a more efficient manner than traditional fuzzy sets, classical relational databases may also be extended to represent and deal with uncertain data with the concept of vague set theory. The extended database model is then called a vague relational database model. Next, we have started to design a vague relational database model. In **chapter 5**, we have defined a new kind of vague functional dependency (called α **-vfd)** based on the notion of α -equality of tuples and the idea of similarity measure of vague sets. Next, we present a set of sound vague inference rules which are similar to Armstrong's axioms for the classical case. Finally, partial α **-vfd** and vague key have been studied with the new notion of α **-vfd** and also tested with examples.

Chapter 6 introduces a new definition of vague multivalued dependency, called α **-vmvd**, on the basis of α -equality of tuples as defined in chapter 5. The definition is shown to be consistent and finally a set of sound and complete inference axioms have been designed and verified for the α **-vmvd**.

Like integrity constraints, normalization process also plays a vital role in designing a good relational database. Thus, as we frame a vague database model, it is essential for the designer to guarantee that data redundancy and anomalies have been minimized. To achieve this, in **Chapter 7**, we have introduced a number of vague normal forms such as vague first (**V1NF**), vague second (**V2NF**), vague third (**V3NF**) and vague Boyce Codd (**VBCNF**) normal forms based on our α **-vfd**. Here we have also presented algorithm for finding vague closure of attribute set which helps to find vague key and plays an important role in the normalization process. Further, since the process of normalization actually uses the idea of decomposition to obtain the desired normal form, one should also confirm the dependency preservation and lossless join properties of decomposition for designing a good database model. Thus in the same chapter we have also dealt with these issues and have designed two algorithms that guarantee that the synthesized relation schemes have the lossless join property and preserve data dependencies in each individual relation.

In **chapter 8**, we have made an attempt to make a theoretical comparison between fuzzy sets and vague sets in processing uncertain queries. Firstly, we have designed an architecture to test uncertain queries. Next, we have presented an algorithm to retrieve membership values for imprecise data represented by fuzzy or vague sets. A similarity measure formula is then used to get each tuple's similarity value with the uncertain query for both fuzzy and vague sets. Finally, the decision maker will supply a threshold value or α-cut based on which a corresponding SQL statement is generated for the given uncertain query. This SQL retrieves different result sets from the database for fuzzy data and vague data. In the present study, we have considered an Employee database and processed some uncertain queries using fuzzy data as well as vague data. Each time it has been observed that our proposed Vague Relational Database Model gives better result than the existing Fuzzy Database Model.

Concluding remarks appear in **chapter 9.** In this chapter, the major contributions of present research have been highlighted. Future scope of work in this domain and possible application areas have also been pointed out in the same chapter.

5. Conclusion

Uncertain data arise constantly in real life from human thought and cognition processes. Thus, clearly, there is always a need to develop a database system which has the capability to treat both precise and imprecise information. In the present thesis, an attempt has been made to provide a complete and unified approach towards the development of a vague relational database model. For a successful blending of vague set theory and relational database, we have extended some of the widely investigated results in classical relational database literature to the fuzzy paradigm and subsequently to vague domain. Our investigation shows that the vague relational database model actually retrieves more accurate result compared to its fuzzy counterpart while processing uncertain queries. This confirms the fact that a vague set has an extra edge over fuzzy set. Hence a vague relational database model, as designed in the present work, may be more fruitful in processing real life data and queries than the conventional fuzzy data models. A DBMS that implements this vague set theoretic concept can thus become a more powerful software product than those currently available.

The **major contributions** of the research work presented in this thesis may be identified as follows:

1. A complete design theory for a vague relational database model has been presented which has not yet been reported in literature. In this quest, new kind of vague data dependencies, namely, vague functional dependency (α**-vfd**) and vague multivalued

dependency (α**-vmvd**) have been defined. In order to make the designed database free from data redundancy and anomalies, several vague normal forms, viz., **VINF**, **V2NF**, **V3NF** and **VBCNF** have been introduced. Algorithms have also been designed that confirm dependency preservation and lossless join decomposition of an unnormalized relation into desired normal form.

2. In the fuzzy framework, a new concept of fuzzy multivalued dependency (α **-fmvd)** has been defined with the idea of α-equality of tuples. This new definition of α**-fmvd** provides a very easy and straightforward way of extending **mvd** to **fmvd** for a fuzzy relational database model and thus it differs from the existing definitions in literature.
3. An architecture has been designed that can process any uncertain query using both fuzzy and vague sets. The present study confirms that a vague relational database model may be more useful in processing uncertain queries than the fuzzy model. Also an algorithm has been proposed that uses a unique concept of attribute independent membership function of any fuzzy set to calculate membership value for the domain values of any fuzzy attribute.

6. Publications Based on the Present Research Work

[1] Mishra J. and Ghosh S., "A Study of Fuzzy Relational Database", **International Journal of Computational Cognition**, Vol. 6, No. 4, pp. 45-50, **2008**.

[2] Mishra J. and Ghosh S., "A Multivalued Integrity Constraint in Fuzzy Relational Database", **International Journal of Computational Cognition**, Vol. 9, No. 2, pp. 72-78, **2011**.

[3] Mishra J. and Ghosh S., "A New Functional Dependency in a Vague Relational Database Model", **International Journal of Computer Applications**, Vol. 39, No. 8, pp. 29-36, **2012**.

[4] Mishra J. and Ghosh S., "Normalization in a Fuzzy Relational Database Model", **International Journal of Computer Engineering & Technology**, Vol. 3, No. 2, pp. 506-517, **2012**.

[5] Mishra J. and Ghosh S., "Vague Normalization in a Relational Database Model", Proceedings of International Conference on Database Theory and Application (**DTA 2012**): Computer Communications in Computer and Information Science (**Springer**), Vol. 352, pp. 332-344, **2012**.

[6] Mishra J. and Ghosh S., " Uncertain Query Processing using Vague Sets : A Better Approach", "3rd International Conference on Computational Intelligence and Information Technology- **CIIT 2013**", IET, pp. 52-60, (**IEEE Xplore**), **2013**.

[7] Mishra J. and Ghosh S., "A Vague Multivalued Data Dependency", **An International Journal of Fuzzy Information and Engineering**, **Springer**, Vol. 5, No. 4, pp. 459-473, **2013**.

[8] Mishra J. and Ghosh S., "Uncertain Query Processing using Vague Sets or Fuzzy Set: Which One is Better?", **International Journal of Computers, Communications and Control (IJCCC)**, Vol. 9, No. 6, pp 730-740, **2014**. (**SCI** journal, **IF** : 0.694)

[9] Mishra J. and Ghosh S., "Dependency Preservation and Lossless Join Decomposition in Vague Normalization", Communicated.

7. References

[1] Al-Hamouz S., Biswas R., "Fuzzy Functional Dependencies in Relational Databases", International Journal of Computational Cognition, Vol. 4, No.1, pp. 39-43, 2006.

[2] Bahar O. and Yazici A., "Normalization and Lossless Join Decomposition of Similarity-Based Fuzzy Relational Databases", International Journal of Intelligent Systems, Vol. 19, No. 10, pp. 885-917, 2004.

[3] Codd E. F., "A Relational Model for Large Shared Data Banks", Comm. of ACM, Vol. 13, No. 6, pp. 377-387, 1970.

[4] Gau W. L. and D. J. Buehrer, "Vague Sets", IEEE Trans. Syst. Man, Cybernetics, Vol. 23, No. 2, pp.610-614, 1993.

[5] Lu A. and Ng W., "Managing Merged Data by Vague Functional Dependencies" Springer-Verlag Berlin Heidelberg LNCS 3288, pp. 259-272, 2004.

[6] Lu A. and Ng W., "Maintaining consistency of vague databases using data dependencies", Data and Knowledge Engineering, Vol. 68, No. 7, pp. 622-641, 2009.

[7] Nakata M., "Dependencies in fuzzy database: Multivalued Dependency", IEEE Int. Conf. on Fuzzy Systems, pp. 1591-1597, 1996.

[8] Raju K.V.S.V.N. and Majumdar A.K., "Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database system", ACM Transactions on Database Systems, Vol. 13, No. 2, pp. 129-166, 1988.

[9] Sözat M. I. and Yazici A. "A complete axiomatization for fuzzy functional and multivalued dependencies in fuzzy database relations", Fuzzy Sets and Systems, Vol. 117, No. 2, pp.161-181, 2001.

[10] Takahashi Y., "Fuzzy database query languages and their relational completeness theorem", IEEE Transactions on Knowledge and Data Engineering, Vol. 5, pp. 122-125, 1993.

[11] Zadeh L. A., "Fuzzy Sets", Information and Control, Vol. 8, No. 3, pp. 338-353, 1965.

[12] Zhao F. and Ma Z. M., "Functional Dependencies in Vague Relational Databases", IEEE International Conference on Systems, Man and Cybernetics, Vol. 5, pp. 4006-4010, 2006.

[13] Zhao F., Ma Z. M. and Yan L., "A Vague Relational Model and Algebra", Fourth International Conference on Fuzzy Systems and Knowledge Discovery (**FSKD 2007**), Vol. 1, pp. 81-85, 2007.

[14] Zhao F. and Ma Z. M., "Vague Query Based on Vague Relational Model", AISC 61, pp. 229-238, Springer-Verlag Berlin Heidelberg, 2009.

Signature of the Candidate

Forwarded by Supervisor:

(Dr. Sharmistha Ghosh)
Professor & Division Chair
Division of Applied Mathematics
School of Basic and Applied Sciences
Galgotias University, Greater Noida
Uttar Pradesh-203201