EasyChair Preprint

№ 11761

Fuzzy Data Mining and Fuzzy Reasoning

Poli Venkatasubba Reddy

January 13, 2024

Fuzzy Data Mining and Fuzzy Reasoning

P. Venkata Subba Reddy
Department of Computer Science and Engineering,
College of Engineering,
Sri Venkateswara University,
Tirupati-517502, India.

ABSTRACT

In this paper, logical data independence for physical data independence for data mining is discussed. The fuzzy databases are studied for data mining with logical and physical data independence. Searching and statistical quantification with fuzzy logic is studied for fuzzy data mining. The fuzzy reasoning for data mining is discussed. The SQL queries are studied for fuzzy data mining.

Keywords

Fuzzy logic, fuzzy reasoning, data mining, fuzzy data mining, SQL

1. INTRODUCTION

The knowledge discovery with data mining necessary to improve business and reduce the cost of business. The data warehousing is necessary for data mining. The information stored in data warehousing may be data or text. The logical and physical data independance are required during join or decomposition of databases. Logical independence is most important before going to physical independence. Various query languages used to mining the information with data or text like SQL, QUEL and Xquery with warehousing. The data mining goal is search and statistical measures an algorithm running on data mining is scalable. It is continuum of analysis and exploration of data or text.

The data mining has different models like frequent item sets, associations, clustering and classification. It is great difficulty data mining with data warehousing for statistical analysis. The fuzzy databases will minimize the statistical analysis during online analytic process (OLAP) or online transaction process (OLTP). The fuzzy data mining methods can be defined using with SQL, QUEL and r Query.

In the following, the logical and phys data independence is studied for data mining by taking relational databases. The logical and physical data independence is studied for fuzzy data mining. The fuzzy reasoning is studied for data mining. The SQL query examples are given for data mining and fuzzy data mining. The fuzzy databases are studied with rough sets

2. RELATIONAL DATABASE

The relational database is defined as Cartesian product of Domains A_1, A_2, .., A_m and is represented as

$R=A_1XA_2X\ldots XA_m$

$t_i=a_{i1}xa_{i2}x,\ldots, x_{aim}$, i=1,…,n are tupes

or

$R(A_1,A_2,\ldots, A_n)$, R is relation, A_1, A_2,… A_m are domains

R(ai1,ai2…, aim, i=1,…,n are tuplesl

For instance,

Cno	Ino	Iname	price
C101	I105	coffee	100
C101	I107	Milk	50
C103	I104	tea	80
C102	I107	milk	80
C101	I108	Sugar	60
C102	I105	coffee	100

consider Price relational database

ino	Iname	price
I105	Coffee	100
I107	Milk	50
I104	Tea	80
i108	Sugar	60
I109	coffee	100

Consider Demand relational database

ino	Iname	Demand
I105	Coffee	80
I107	Milk	60
I104	Tea	100
I108	Sugar	50
I109	coffee	80

Lossless Join is given by

Cno	Ino	Iname	price
C101	I105	coffee	100
C101	I107	Milk	50
C103	I104	tea	80
C102	I107	milk	80
C101	I108	Sugar	60
C102	I105	coffee	100

Lossless decomposition is given by

ino	Iname	price
I105	Coffee	100
I107	Milk	50
I104	Tea	80
108	Sugar	60
I105	Coffee	100

ino	Iname	Demand
I105	Coffee	80
I107	Milk	60
I104	Tea	100
i108	Sugar	50
I109	Coffee	80

3. DATA WHEREHOUSING AND DATA MINING

3.1 Data warehousing

The data mining can be performed using data warehousing. In data warehousing information is stored with data or text. The Relational database with data or data as text used to keep the information. For example, super market and new paper business data warehousing [4]

Consider Purchase database

Cno	Ino	Iname	price
C101	I105	coffee	100
C101	I107	Milk	50
C103	I104	tea	80
C102	I107	milk	80
C101	I108	Sugar	60
C102	I105	coffee	100

3.2 Data Mining

Data mining is knowledge discovery process dealing with methods like frequent items, association rules, clustering records, representation of tree, classification of trees and uncertainty in data[2,4].

In the following some of the methods are discussed

Customers who purchased more than one Item. Is given by

Cno	Ino	Iname	price
C101	I105	coffee	100
C101	I107	Milk	50
C103	I104	tea	80
C102	I107	milk	80
C101	I108	Sugar	60
C102	I105	coffee	100

SELECT P.Cno, P.Ino, Iname, COUNT(*)

FROM purchase P

WHERE COUNT(*)>1.

Cno	Ino	COUNT
C101	I105	2
C102	I105	2

Association rules.

Customers who purchased coffee and milk. is given by

<coffee.=><milk>

SELECT P.Cno, P.Ino

FROM purchase P

WHERE Iname=”coffee” and Iname=”milk”

Cno	Ino
C101	I105
C102	I105

4. FUZZY DATA MINING

Fuzzy Data Mining is knowledge discovery process with data associated with uncertainty or incompleteness. The fuzzy logic[7, 8] is more suitable to deal with such data because fuzzy logic deals with commonsense rather than likelihood.

Fizzy Relational Databases are discussed with Rough set theory. Rough Set theory is another approach to incomplete information[16]. The incomplete Information may be deal with fuzzy logic.

Definition: Given some universe of discourse X, a fuzzy rough set is defined as pair $\{t, \mu_d(t)\}$, where d is domains and membership function $\mu_d(x)$ taking values on the unit interval[0,1] i.e. $\mu_d(t) \rightarrow [0,1]$, where $t_i \epsilon X$ is tuples .

	d_1	2_2	.	d_m	μ
t_1	a_{11}	a_{12}	.	a_{1m}	$\mu_d(t_1)$
t_2	a_{21}	a_{22}		A_{2m}	$\mu_d(t_2)$
.	.	.	.	.	.
t_n	a_{1n}	a_{1n}	.	A_{nm}	$\mu_d(t_n)$

Let C and D be the fuzzy rough sets .

The operations on fuzzy rough sets are given as

$1\text{-}C = 1 - \mu_C(x)$ Negation

$CVD = \max\{\mu_C(x), \mu_D(x)\}$ Union

$C \Lambda D = \min\{\mu_C(x), \mu_D(x)\}$ Intersection

Consider Price fuzzy database

ino	Iname	price	μ
I105	Coffee	100	0.8
I107	Milk	50	0.4
I104	Tea	80	0.7
i108	Sugar	60	0.5
I109	coffee	100	0.8

The Negation of Price fuzzy database is given by

ino	Iname	price	μ
I105	Coffee	100	0.2
I107	Milk	50	0.5
I104	Tea	80	0.8
i108	Sugar	60	0.5
I109	coffee	100	0.2

Consider Demand fuzzy relational database

ino	Iname	Demand	μ
I105	Coffee	80	0.7
I107	Milk	60	0.5
I104	Tea	100	0.8
I108	Sugar	50	0.4
I109	coffee	80	0.7

Fuzzy lossless Natural Join of Demand and Price is Union and given as

ino	Iname	Demand	price	μ
I105	Coffee	80	100	0.8
I107	Milk	60	50	0.5
I104	Tea	100	80	0.8
I108	Sugar	50	60	0.5
I109	Coffee	80	100	0.8

Fuzzy intersect of Demand and Price is given by

ino	Iname	Demand	price	μ
I105	Coffee	80	100	0.7
I107	Milk	60	50	0.4
I104	Tea	100	80	0.7
I108	Sugar	50	60	0.4
I109	Coffee	80	100	0.7

Fuzzy lossless decomposition is given by

ino	Iname	Demand	μ
I105	Coffee	80	0.8
I107	Milk	60	0.5
I104	Tea	100	0.8
i108	Sugar	50	0.5
I109	Coffee	80	0.8

ino	Iname	price	μ
I105	Coffee	100	0.8
I107	Milk	50	0.5
I104	Tea	80	0.8
108	Sugar	60	0.5
I105	Coffee	100	0.8

For instance fuzzy software database , consider best software Company is given by

Company	μ
IBM	0.8
Microsoft	0.9
Novel	0.6
Oracle	0.7
Google	0.75

Rich software company with fuzziness >.7 is given by

SELECT P.software

FROM software P

WHERE μ>0.7

Company	μ
IBM	0.8
Microsoft	0.9
Google	0.75

Consider fuzzy database

Cno	Ino	Iname	price	μ
C101	I105	coffee	100	0.8
C101	I107	Milk	50	0.5
C103	I104	tea	80	0.8
C102	I107	milk	80	0.5
C101	I108	Sugar	60	0.8
C102	I105	coffee	100	0.8

Fuzzy Frequency Items

Customers who purchased more than one Item. Is given by

SELECT P.Cno, P.Ino, Iname, COUNT(*)

FROM purchase P

WHERE μ>0.6.

Cno	Ino	COUNT
C101	I105	2
C102	I105	2

Fuzzy Association rules.

Customers who purchased coffee and milk is given by

<coffee.=><milk>

SELECT P.Cno, P.Ino

FROM purchase P

WHERE Iname="coffee" and Iname="milk" and μ>.7

Cno	Ino
C101	I105

Fuzzy functional dependency and Association rule

Let R is Relational Database. T is set of tuples. The Functional dependency

FD:X→Y exits if t1[X]=t2[X] then t1[Y]=t2[Y]

A Fuzzy Functional Dependency is defend by

FD:X→Y exits if t1[X] .approximately equal. t2[X] then t1[Y]. approximately equal. t2[Y]

i.e.

if t1[X].EQ.t2[X] then t1[Y].EQ.t2[Y]

The fuzzy conditional inference for FFD using Gödel's definition is given by

FFD:X→Y=

1 μ(t1) <=μ(t2)

μ (t1) μ(t1) >μ(t2)

Consider the fuzzy database

Cno	Ino	Iname	price	μ
C101	I105	coffee	100	0.8
C101	I107	Milk	50	0.5
C103	I104	tea	80	0.8
C102	I107	milk	80	0.5
C101	I108	Sugar	60	0.8
C102	I105	coffee	100	0.8

The fuzzy equality Cno is given by

Cno	EQ
<C101><=><C101>	1
<C102><=><C102>	1
<C103><=><C103>	1

The fuzzy equality Iname given by

Iname	EQ
<Coffee><=><Milk>	0.8
<Milk><=><Coffee2>	0.8
<Tea><=><Sugar>	0.4

The FFD:Cno→Iname is given by

Cno	Iname	EQ
<C101><=><C101>	<Coffee><=><Milk>	0.8
<C102><=><C102>	<Milk><=><Coffee2>	0.8
<C103><=><C103>	<Tea><=><Sugar>	0.4

The fuzzy association may be given for items with EQ>o.5

Cno	Iname	EQ
<C101><=><C101>	<Coffee><=><Milk>	0.8
<C102><=><C102>	<Milk><=><Coffee2>	0.8

.

5. REASONING WITH FUZZY DATA MINING

The Business Intelligence may be studied with Fuzzy data Mining.

Consider the fuzzy conditional inference

If x is A then x is B

Muzimoto fuzzy reasoning is given by [3]

If x is A then x is B

X is more A

X is more B

Provided $\mu_A(x) < \mu_B(x)$

If x is Demand then x is price

X is more demand

X is more Price

From the above tables, more Price is given by

ino	Price
I105	0.89
I107	0.71
I104	0.84
i108	0.71
I109	0.89

Find the Items with Price >=0.8

SELECT ino

FROM Items

Where Price >=0.8

ino
I105
I104
I109

Consider the Iname with fuzzy frequency algorithm.

The fuzzy frequency algorithm is replacing MapReducing algorithm. finding frequency data sets with different fuzzy frequency.

Cno	Ino	Iname	frequency
C101	I105	coffee	0.75
C101	I107	Milk	0.5
C103	I104	tea	0.7
C102	I107	milk	0.4
C101	I108	Sugar	0.3
C102	I105	coffee	0.0.8

item sets for above fuzzy databases is given

Cno	Ino	Iname	frequency
C101, C102	I105	coffee	0.75
C101, C102	I107	Milk	0.4
C103	I104	tea	0.7
C101	I108	Sugar	0.3

Find the frequency of data sets >=0.6

SELECT *

FROM Items

Where frequency>=0.6

Cno	Ino	Iname	frequency
C101, C102	I105	coffee	0.75
C103	I104	tea	0.7

6. CONCLUSION

Data Mining is Knowledge Discovery Process with data warehousing. The logical and physical data independent is studied by taking examples. The fuzzy reasoning is studied with Fuzzy Data Mining. The advantages of fuzzy Data Mining are minimizing data aces costs of Big Data, minimize the multiple copies of functions and objects. Fuzzy databases data sets replace the any MapReducing algorithms. In similar line Xquery may be studied.

7. ACKNOWLEDGMENTS

The author thanks referees for their valuable suggestions

8. REFERENCES

[1] Billey P. Buckless, Fuzzy Databases and Their Applications, Fuzzy Information and Decision Process, M.M.Gupta and E.Sanchez (ed), 1982

[2] Micheline Kamber, and Jian Pei. *Data mining:* Han, Jiawei *concepts and techniques*. Morgan kaufmann, 2006

[3] S. Fukami,M. Muzimoto,K.Tanaka,Some Considerations on Fuzzy conditional Inference, Fuzzy Sets and Systems, 4 (1980) 243-278.

[4] R.ramakrishnan, J. Gehrike, Database Management Systems, McGrahill, 2003

[5] J.D. Ullman, Principles of Database Systems, Galgotia Publications, , 1999.

[6] P. Venkata Subba Reddy and M.Shyam Babu. , " Some methods of reasoning for conditional propositions ", fuzzy sets and systems , 52,(1992)229-250..

[7] L.A. Zadeh, "Calculus of Fuzzy restrictions", Fuzzy sets and their applications to cognitive and decision processes, L.A.Zadeh,K.S.Fu,M.Shimura,Eds,New York, Academic, (1975,)1-39.

[8] L.A Zadeh," Fuzzy sets", Information Control,.8(1965)38-353.