FuzzMapReduceAL: Fuzzy MapReduce Algorithmic Language

Venkata Subba Reddy Poli

March 5, 2023

FuzzMapReduceAL: Fuzzy MapReduce Algorithmic Language

Poli Venkata Subba Reddy
Department of Computer Science and Engineering
College of Engineering
Sri VenkateswaraUniversity
Tirupati, India
vsrpoli@hotmail.com

***Abstract*—Fuzzy Data Mining (FDM) is knowledge discovery process for easy understanding , reasoning decision making not only for incomplete information but also complete information. The fuzzy MapReduce algorithmic language is needed to design the MapReduce algorithm for fuzzy data mining methods and classifications. In this paper, fuzzy MapReduce algorithmic language is studied for Data Mining. The data mining methods and classifications are brought under FuzzMapReduceAL. The business intelligence is given as an example.**

***Keywords*— fuzzy logic, fuzzy database, fuzzy data mining, fuzzy MapReduce algorithms**

I. Introduction

Zadeh [15] has introduced fuzzy set as a model to deal with imprecise, inconsistent and inexact, vague and approximate information. The fuzzy set is a class of objects with a continuum of grades of membership.

It is necessary to discuss incomplete information with fuzzy logic.

The fuzzy set A of X is characterized as its membership function A = $\mu_A(x_j)$ and ranging values in the unit interval [0, 1]

$\mu_A(x_j)$: X →[0, 1], x∈ X, where X is Universe of discourse.

A = $\mu_A(x_1)/x_1$ + $\mu_A(x_2)/x_2$ + … + $\mu_A(x_n)/x_n$, "+" is union

For instance, the fuzzy proposition "x is best car"

Sales={0.5/Suzuki+0.7/Skoda 0.9/Benz +0.8/Toyota + 0.6/Honda }

II. FUZZY MAPREDCE ALGORITHMIC LANGUAGE

Fuzzy Data mining is knowledge discovery process from Fuzzy databases. some of the methods are fuzzy frequent items. Fuzzy association rules. Fuzzy clustering and classifications like fuzzy reasoning.

The fuzzy algorithmic language is necessary to study for fuzzy data mining

1. BEGIN initial
   END terminal
2. input fuzzy database
   output database
3. read fuzzy database
4. write fuzzy variables
5. fuzzy statement
6. fuzzy Negation
   1-A
7. Fuzzy Union
   C = AUB
8. Fuzzy Intersection
   C = A∩B
9. Fuzzy join
   C=A⋈ B
10. Fuzzy decompositions
    C=A, B
11. Fuzzy frequency
    Selection $\sigma_{R=\alpha}$
12. Fuzzy clustering
    R=$A_1, A_2, \ldots A_n$
13. Fuzzy Association
    A⇔B
14. Projection
15. Proj A = min{$\mu_A(x_1)/x_1$, $\mu_A(x_2)/x_2$, … , $\mu_A(x_n)/x_n$ }
16. fuzzy inference
    if x is A_i then y is B_i
17. Fuzzy composition
    R=A1oA1→B
18. return A

III. FUZZY MapReduce

Fuzzy Data Mining is knowledge discovery process with data associated with uncertainty or incomplete information. Fuzzy MapReducing may be studied with methods and classifications.

The fuzzy MapReduce algorithms two functions Mapping and Reducing the map data.

Fuzzy membership function $\mu_d(x)$ taking values on the unit interval[0. 1] i.e. $\mu_d(t)$→[0. 1]. where t_i∈X is tuples .

TABLE I. Fuzzy data set

	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}	$\mu_d(t_n)$

$\mu_D(r)= \mu_d(t_1) + \mu_d(t_2) + \ldots + \mu_d(t_n)$, Where "+" is union, D is domain and t_i are tupls..

TABLE II. Price relational data sets

Cno	Ino	Iname	price
C101	I105	Shirt	70
C101	I107	Dress	50
C103	I104	Pants	60
C102	I107	Dress	50
C101	I108	Jacket	55
C102	I105	Shirt	70

TABLE III. Fuzzy relational data set

Cno	Ino	Iname	price
C101	I105	Shirt	0.7
C101	I107	Dress	0.5
C103	I104	Pants	0.6
C102	I107	Dress	0.5
C101	I108	Jacket	0.6
C102	I105	Shirt	0.7

A. Negation

TABLE IV. Negation of Price

Cno	Ino	Iname	Negation of price
C101	I105	Shirt	0.9
C101	I107	Dress	0.4
C103	I104	Pants	0.7
C102	I107	Dress	0.4
C101	I108	Jacket	0.5
C102	I105	Shirt	0.9

TABLE V. Sales

Cno	Ino	Iname	sales
C101	I105	Shirt	20
C101	I107	Dress	10
C103	I104	Pants	16
C102	I107	Dress	14
C101	I108	Jacket	12
C102	I105	Shirt	18

The Mapping TABLE IV Reduce to TABLE V by fuzzification

TABLE VI. Fuzzy relational data set

Cno	Ino	Iname	sales
C101	I105	Shirt	0.8
C101	I107	Dress	0.4
C103	I104	Pants	0.6
C102	I107	Dress	0.5
C101	I108	Jacket	0.5
C102	I105	Shirt	0.7

B. Union

TABLE VII. Sales U Price

Cno	Ino	Iname	Sales U price
C101	I105	Shirt	0.8
C101	I107	Dress	0.5
C103	I104	Pants	0.6
C102	I107	Dress	0.5
C101	I108	Jacket	0.6
C102	I105	Shirt	0.7

C. Intersection

TABLE VIII Sales ∩ Price

Cno	Ino	Iname	Sales ∩ price
C101	I105	Shirt	0.7
C101	I107	Dress	0.4
C103	I104	Pants	0.6
C102	I107	Dress	0.5
C101	I108	Jacket	0.5
C102	I105	Shirt	0.7

D. Implication

TABLE IX. Sales→Price

Cno	Ino	Iname	Sales→price
C101	I105	Shirt	0.9

C101	I107	Dress	1.0
C103	I104	Pants	0.9
C102	I107	Dress	1.0
C101	I108	Jacket	1.0
C102	I105	Shirt	1.0

E. Fuzzy frequency

Fuzzy frequency- = 0.1/1+0.3/2+0./3+0.6/4+0.7/7

TABLE X. fuzzy frequency

Cno	frequency
C101	0.5
C102	0.3
C103	0.1

F. Fuzzy Clustering

Cluster with fuzziness>0.5 and <=0.5,

TABLE XI. fuzzy clustering

Cno	Ino	Iname	salesVprice
C101 C101	I105	Shirt	0.8
	I108	Jacket	0.5
C102 C102	I105	Shirt	0.7
	I107	Dress	0.5
C103	I104	Pants	0.6

G. Fuzzy Association

If EQ(t1(X),t2(X)) then EQ(t1(Y) ,t2(Y))
EQ(t1(X),t2(X)) →EQ(t1(Y) ,t2(Y))
=min{ EQ(t1(X),t2(X)) , EQ(t1(Y) ,t2(Y))}
= min{ 1, EQ(t1(Y) ,t2(Y))}
=,EQ(t1(Y) ,t2(Y))

The fuzzy equivalence is defined by
$\mu_{EQ(t1(Y),t2(Y))}(Y) = \min\{\mu_{t1}(y), \mu_{t2}(y)\}$

The fuzzy association dependency (FAD) "⇔" may be give as

TABLE XII. Association

Cno	Ino	Iname	sales
C101	I105⇔I107	Shirt⇔Dress	0.4
C103	I104	Pants	0.6
C102	I107⇔I105	Dress⇔Shirt	0.5

TABLE XIII. Fuzzy relational sales data set.

Cno	Ino	Iname	sales
C101	I105	Shirt	0.8
C101	I107	Dress	0.4
C103	I104	Pants	0.6
C102	I107	Dress	0.5
C101	I108	Jacket	0.5
C102	I105	Shirt	0.7

The fuzzy multivalve association may e defined as is defined as

If EQ(t1(X),t2(X),t3(X)) then EQ(t1(Y) ,t2(Y)) or EQ(t2(Y) ,t3(Y)) or EQ(t1(Y) ,t3(Y))
= min{EQ(t1(X),t2(X),t3(X)), min(EQ(t1(Y) ,t2(Y)) , EQ(t2(Y) ,t3(Y)) , EQ(t1(Y) ,t3(Y)))}
= min{1, min(min (μ_{t1}(Y), μt_2(Y)) , min (μ_{t2}(Y), $\mu t3$(Y)), min (μ_{t1}(Y), $\mu t3$(Y))}
The FAMVD is FAD.

The fuzzy association ⇔ may be give as,using AFMVD

A. Natural Join

sales ⋈ price=min{ sales, price)

TABLE XIV. Sales ⋈ Price

Cno	Ino	Iname	sales
C101	I105⇔I107 ⇔I108	Shirt⇔Dress ⇔Jacket	0.8 0.4 0.5
C103	I104	Pants	0.6
C102	I107⇔I105	Dress⇔Shirt	0.5 0.7

B. Normalization

Using table 10, the normal forms are given by

TABLE XV. Sales

Cno	Ino	Iname	sales
C101	I105	Shirt	0.8

C101	I107	Dress	0.5
C103	I104	Pants	0.6
C102	I107	Dress	0.5
C101	I108	Jacket	0.6
C102	I105	Shirt	0.7

TABLE XVI.Price

Cno	Ino	Iname	price
C101	I105	Shirt	0.8
C101	I107	Dress	0.5
C103	I104	Pants	0.6
C102	I107	Dress	0.5
C101	I108	Jacket	0.6
C102	I105	Shirt	0.7

IV. FUZZY K-MEANS ALGORITHM

The fuzzy k-means data set algorithm (FKCA) is optimization algorithm for fuzzy data sets

```
best=k-means( k-fuzzy data sets)
for in range(1,k)
C=fuzzy-association
if   k-means( k-fuzzy data sets)<best
best=C
return best
```

for example

consider sorted fuzzy sets of TABLE V is given by

TABLE XVII.. Sorted fuzzy data sets

Cno	Ino	Iname	sales
C101	I105	Shirt	0.8
C101	I107	Dress	0.4
C101	I108	Jacket	0.5
C102	I107	Dress	0.5
C102	I105	Shirt	0.7
C103	I104	Pants	0.6

Apply FAD 1^{st} iteration

TABLE XVIII. First iteration

Cno	Ino	Iname	sales
C101	I105⇔I107	Shirt⇔Dress	0.4
C101	I108	Jacket	0.5
C102	I107	Dress	0.5
C102	I105	Shirt	0.7
C103	I104	Pants	0.6

Similarly continue do iteration, the optimization fuzzy data sets is given by

TABLE XIX. Optimization data sets

Cno	Ino	Iname	sales
C101	I105⇔I107 ⇔I108	Shirt⇔Dress ⇔Jacket	0.4
C103	I104	Pants	0.6
C102	I107⇔I105	Dress⇔Shirt	0.5

V. FUZZY DATA MINING REASONING

The fuzzy reasoning may be applied for Fuzzy data Mining.

Consider the more Demand fuzzy database by decomposition

The fuzzy rata mining reasoning may be performed using Zadeh [14] fuzzy conditional inference

Zadeh fuzzy inference is given when consequent part is not known as

$= \min(1, 1-\min(A_1, A_2 ,\dots ,A_n)+ B)$

Mamdani[2] inference is given as

if x_1 is A_1 and x_2 is A_2 and … and x_n is A_n then y is B

$= \min(A_1, A_2 ,\dots ,A_n , B)$

The fuzzy conditional inference may be derived from precedent part.

if x_1 is A_1 and x_2 is A_2 and … and x_n is A_n then y is B

$= \min(A_1, A_2 ,\dots ,A_n , B)$

$=\min(A_1, A_2 ,\dots ,A_n ,1)$, where B=1 because B is not known.

$=\min(A_1, A_2 ,\dots ,A_n ,1)$

$\min(A_1, A_2 ,\dots ,A_n)$

The fuzzy inference may be derived fro composition.

$R=A1o(A\rightarrow B)$

If x is sales then x is price

x is more sales

----

x is more sales o (sales→price)

Zadeh [14] fuzzy reasoning is given by

x is more sales o (min{1, 1-sales+price})

Momdani[1] fuzzy reasoning is given by
x is more sales o (min{sales, Price})

Prosed fuzzy reasoning is given by
If x is sales then x is price
x is more sales
----
x is more sales o(sales)

TABLE XX. More sales

Cno	Iname	μ
C101	Shirt	0.89
C101	Dress	0.77
C103	Pants	0.94
C102	Dress	0.70
C101	Jacket	0.89
C102	Shirt	0.63

TABLE XXI Fuzzy reasoning

Cno	Iname	Zadeh	Mamdani	Reddy
C101	Shirt	0.89	0.8	0.8
C101	Dress	0.77	0.5	0.6
C103	Pants	0.94	0.8	0.9
C102	Dress	0.70	0.5	0.5
C101	Jacket	0.89	0.6	0.8
C102	Shirt	0.63	0.4	0.4

ACKNOWLEDGMENT

The author thank to iFuzzy2020 Conference Chair for accepting this paper.

REFERENCES

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

[2] E.H. Mamdani and S. Assilian. An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller, Int. J. Man-Machine Studies, vol.7,pp.1-13, 1975,

[3] Daniel Sánchez Alvarez and Antonio F. Gómez Skarmeta , "A fuzzy Language", fuzzy Sets and Systems,.vol. 141,.no.3,pp.335-390, 2014

[4] W. Pedrycz and F. Gomide, Introduction to fuzzy Sets, Cambridge, MIT, 1998.

[5] E. Santos , " fuzzy algorithms", Information and Control , 17,(1970).

[6] H.A. Simon , " The structure of ill structured problems" , Artificial Intelligence , vol.4,pp.181-201,1973.

[7] K.Tanaka and Mizumoto , " fuzy Programs and their Executions" , fuzzy sets and their applications to cognitive and decision processes, , L.A. Zadeh , King-Sun FU , KokichiTanaka, Masamichi Shimura, Eds., Academic Press, New York,pp..47-76., 1975

[8] Poli Venkata Subba The proposed "Fuzzy Conditional Inference for Mediiagnosis", Second International Conference on Fuzzy Theory and Technology, Proceedings, Abstract and Summaries of FT&T1993, University of North-Carolina, Duke University, October 13-16, USA,,pp.193-19, 19935.

[9] Poli Venkata Subba Reddy Fuzzy data mining and web intelligence, Fuzzy Theory and Its Applications (iFUZZY), 2015 International Conference on, Year:pp.74 – 79, 2015.

[10] Poli Venkata Subba Reddy, " FUZZYALGOL: Fuzzy Algorithmic Language for Design Fuzzy Algorithms", International Journal of Computer Science and Engineering, 2,2,(2010).

[11] Poli Venkata Subba Reddy, Fuzzy logic based on Belief and Disbelief membership functions, Fuzzy Information and Engineering, vol.,9, no.9,405-422,2017..

[12] L.A. Zadeh, "A note on web intelligence, world knowledge and fuzzy logic", Data and Knowledge Engineering,, vol.50,pp.91-304,2004

[13] L. A Zadeh, "Calculus of Fuzzy Restrictions", In Fuzzy Sets and their Applications to Cognitive and Decision Processes, L. A. Zadeh, King-Sun FU, Kokichi Tanaka and Masamich Shimura (Eds.), Academic Press, New York, pp.1-40, 1975.

[14] L.A.Zadeh , "fuzzy Algorithms" ,Information and Control , vol.12, pp94-104, 1965

[15] L.A.Zadeh , "fuzzy Logic" , IEEE Computer , pp.83- 92,1988.