Fuzzy k-Means Data Mining Association Algorithm

Venkata Subba Reddy Poli

June 7, 2020

Fuzzy k-Means Data Mining Association Algorithm

Poli Venkata Subba Reddy
lDepartment of Computer Science and Engineering
Sei Venkateswara University
Tirupat, India
email:vsrpoli@hotmail.com

***Abstract*— Bayesian theory needs exponential complexity to deal incomplete information. The fuzzy logic deal incomplete information with linear complexity. Fuzzy data mining is knowledge discovery process to deal with incomplete information.. In this paper, fuzzy MapReduce algorithms are studied for Data Mining The fuzzy k-means association algorithm is studied using fuzzy functional dependencies for association rules. The business intelligence is given as an example.**

***Keywords*— fuzzy logic, fuzzy database, fuzzy data mining, fuzzy MapReduce algorithms, fuzzy k-means clustering**

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.

The incomplete information with fuzzy logic is given by.

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 }

Data mining is knowledge discovery process which is very complex. Fuzzy data mining will made easy data mining. The k-means fuzzy algorithm is studied to reduce the complexity of data mining.

II. FUZZY DATA MINING

Zadeh[8] proposed fuzzy logic to define incomplete information. Fuzzy Data Mining is knowledge discovery process with data associated with uncertainty or incomplete information.

Fuzzy data mining methods negation, union, intersection, implication, frequency, clustering and association are useful to knowledge discovery process with inherently defend with fuzziness.

The fuzzy MapReducing algorithms two functions Mapping read fuzzy data sets and Reducing write the after operations.

Definition: Given some universe of discourse X. fuzzy relational data sets are 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 $t_i \in X$ 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..

Let C and D be the fuzzy data sets .

TABLE II. Price relational data sets

Cno	Ino	Iname	price
C101	I105	coffee	70
C101	I107	Milk	50
C103	I104	tea	60
C102	I107	milk	50
C101	I108	Sugar	55
C102	I105	coffee	70

The sale is defined intermittently with fuzziness.

$\mu_{Price}(x)$=0.7/70+0.6/60+0.6/55+0.5/50

or

Fuzziness may be defined with function

$\mu_{Price}(x)= (1+(\text{sales}-100)/100))^{-1}$ price <=100

=1 price >100

The Mapping TABLE 6 and Reduce to TABLE. VII by fuzzification

TABLE III. Fuzzy relational data set

Cno	Ino	Iname	price
C101	I105	coffee	0.7
C101	I107	Milk	0.5
C103	I104	tea	0.6
C102	I107	milk	0.5
C101	I108	Sugar	0.6
C102	I105	coffee	0.7

A. Negation

TABLE 8. Negation of Price

Cno	Ino	Iname	Negation of price
C101	I105	coffee	0.9
C101	I107	Milk	0.4
C103	I104	tea	0.7
C102	I107	milk	0.4
C101	I108	Sugar	0.5
C102	I105	coffee	0.9

TABLE IV. Sales

Cno	Ino	Iname	sales
C101	I105	coffee	20
C101	I107	Milk	10
C103	I104	tea	16
C102	I107	milk	14
C101	I108	Sugar	12
C102	I105	coffee	18

The sale is defined intermittently with fuzziness.

$\mu_{sales}(x)=0.7/70+0.6/60+0.6/55+0.5/50$

or

Fuzziness may be defined with function

$\mu_{sales}(x)= (1+(sales-100)/100))^{-1}$ sales <=100
=1 sales >100

The Mapping TABLE IV Reduce to TABLE V by fuzzification

TABLE V. Fuzzy relational data set

Cno	Ino	Iname	sales
C101	I105	coffee	0.8
C101	I107	Milk	0.4
C103	I104	tea	0.6
C102	I107	milk	0.5
C101	I108	Sugar	0.5
C102	I105	coffee	0.7

B. Union

TABLE VI. Sales U Price

Cno	Ino	Iname	Sales U price
C101	I105	coffee	0.8
C101	I107	Milk	0.5
C103	I104	tea	0.6
C102	I107	milk	0.5
C101	I108	Sugar	0.6
C102	I105	coffee	0.7

C. Intersection

TABLE VII Sales ∩ Price

Cno	Ino	Iname	Sales ∩ price
C101	I105	coffee	0.7
C101	I107	Milk	0.4
C103	I104	tea	0.6
C102	I107	milk	0.5
C101	I108	Sugar	0.5
C102	I105	coffee	0.7

D. Implication

TABLE VIII. Sales→Price

Cno	Ino	Iname	Sales→price
C101	I105	coffee	0.9
C101	I107	Milk	1.0
C103	I104	tea	0.9
C102	I107	milk	1.0
C101	I108	Sugar	1.0
C102	I105	coffee	1.0

E. Fuzzy frequency

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

TABLE IX. 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 X. fuzzy clustering

Cno	Ino	Iname	salesVprice
C101 C101	I105	coffee	0.8
	I108	Sugar	0.5
C102 C102	I105	coffee	0.7
	I107	milk	0.5
C103	I104	tea	0.6

III. FUZZY FUNCTIONAL DEPENDENCY FOR ASSOCIATION

Let R is Relational Data set. t is set of tuples.
The functional dependency FD:X→Y or Y depending on X is defined by
If t1(X)=t2(X) then t1(Y) = t2(Y)
The association property of data mining may be defined with fuzzy functional dependency.
The fuzzy functional dependency FFD;X→Y or Y depending on X is defined by
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
$\mu_{EQ(t1(Y),t2(Y))}(Y) = \min\{\mu_{t1}(y), \mu_{t2}(y)\}$

Consider the TABLE 10 . The fuzzy association dependency (FAD) "⇔" may be give as

TABLE XI. Association

Cno	Ino	Iname	sales
C101	I105⇔I107	Coffee⇔Milk	0.4
C103	I104	tea	0.6
C102	I107⇔I105	Milk⇔coffee	0.5

TABLE XII. Fuzzy relational sales data set.

Cno	Ino	Iname	sales
C101	I105	coffee	0.8
C101	I107	Milk	0.4
C103	I104	tea	0.6
C102	I107	milk	0.5
C101	I108	Sugar	0.5
C102	I105	coffee	0.7

The multivalve dependency(MVD) is defined as
MVD;X~Y or Y is multivalve depending on X is defined by
If t1(X)=t2(X)=t3(X) then t1(Y) = t2(Y) or t2(Y)=t3(Y) or t1(Y)=t3(Y)
The fuzzy association may be defined as multi valued dependency.
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))
The fuzzy association multivalve dependency(FAMVD) may defined by using Mamdani fuzzy conditional inference [3]
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.

Consider the TABLE 17. The fuzzy association ⇔ may be give as
TABLE XIII. Association using AFMVD

A. Natural Join

sales ⋈ price=min{ sales, price)

TABLE IVX. Sales ⋈ Price

Cno	Ino	Iname	sales
C101	I105⇔I107 ⇔I108	Coffee⇔Milk ⇔Sugar	0.8 0.4 0.5
C103	I104	tea	0.6
C102	I107⇔I105	Milk⇔coffee	0.5 0.7

B. Normalization

Using table 10, the normal forms are given by

TABLE VX. Sales

Cno	Ino	Iname	sales
C101	I105	coffee	0.8
C101	I107	Milk	0.5
C103	I104	tea	0.6
C102	I107	milk	0.5
C101	I108	Sugar	0.6
C102	I105	coffee	0.7

TABLE VIX.Price

Cno	Ino	Iname	price
C101	I105	coffee	0.8
C101	I107	Milk	0.5
C103	I104	tea	0.6
C102	I107	milk	0.5
C101	I108	Sugar	0.6
C102	I105	coffee	0.7

IV. FUZZY K-MEANS ASOCIATION THM FOR ASSOCIATION RULES

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

FKCA is given by, using FAD and FAMVD

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 VIIX.. Sorted fuzzy data sets

Cno	Ino	Iname	sales
C101	I105	coffee	0.8
C101	I107	Milk	0.4
C101	I108	Sugar	0.5
C102	I107	milk	0.5
C102	I105	coffee	0.7
C103	I104	tea	0.6

Apply FAD 1st iteration on TABLE VIIX

TABLE VIIIX. First iteration

Cno	Ino	Iname	sales
C101	I105⇔I107	Coffee⇔Milk	0.4
C101	I108	Sugar	0.5
C102	I107	milk	0.5
C102	I105	coffee	0.7
C103	I104	tea	0.6

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

TABLE IXX. Optimization data sets

Cno	Ino	Iname	sales
C101	I105⇔I107 ⇔I108	Coffee⇔Milk ⇔Sugar	0.4
C103	I104	tea	0.6
C102	I107⇔I105	Milk⇔coffee	0.5

”.

ACKNOWLEDGMENT

The author thanks to Sri Venkateswara University for providing faculties to carry out this work .

REFERENCES

[1] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., H. Mamdani and S. Assilian. An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller, Int. J. Man-Machine Studies, vo.7, pp1-13., 1975.

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

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

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

[4] E. Santos , “ fuzzy algorithms”, Information and Control , 17,(1970).

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

[6] 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

[7] P. 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.

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

[9] P. Venkata Subba Reddy, " FUZZYALGOL: Fuzzy Algorithmic Language for Design Fuzzy Algorithms", International Journal of Computer Science and Engineering, 2,2,(2010).
[10] Poli Venkata Subba Reddy, Fuzzy logic based on Belief and Disbelief membership functions, Fuzzy Information and Engineering, vol.,9, no.9,405-422,2017..
[11] L.A. Zadeh, "A note on web intelligence, world knowledge and fuzzy logic", Data and Knowledge Engineering,, vol.50,pp.91-304,2004
[12] 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.
[13] L.A.Zadeh , "fuzzy Algorithms" ,Information and Control , vol.12, pp94-104, 1965
[14] L.A.Zadeh , "fuzzy Logic" , IEEE Computer , pp.83- 92,1988