

# Top-k Ranking with Membership Function for Deductive Database

Keerachart Suksut, Pasapitch Chujai, Nittaya Kerdprasop, and Kittisak Kerdprasop

**Abstract**—The objective of this research is to study top-k ranking in the queries that are ambiguous. In this paper we demonstrate our query answering strategy for ranking world population for deductive data base. The cause of most problems are wrong ranking because some questions have ambiguity such as “Find the country which have the population between 1,500,000 and 3,000,000 people by the most densely population is approximately 2,400,000 people.” This research proposes top-k ranking technique using membership function to evaluate and rank possible answers. We show comparative results for each kind of membership function.

**Index Terms**—ER data mining, datalogtop-k, datalog, deductive database.

## I. INTRODUCTION

Ranking of data is highly important because its help to define direct scope of target with goods which we want to attribute or design and create new products which market demand during that time [1].

Ranking can make by determination only one condition or more. Ranking the fuzzy data for finding least fuzzy answer by using membership function technique to reduce fuzzy of data. Result ranking the fuzzy data have more precision. Correct ranking can increase business profit highly [2].

In paper, main idea is apply the membership function to reduce fuzzy data in part of membership value computation. We show some difference of apply the membership function (Triangular Function, Trapezoidal Function, Left Shoulder Function and Right Shoulder Function).

Aforementioned, researcher exhibit ranging technique of the fuzzy data by member function for find the best ranking of data to increase more accuracy and precise.

## II. BASIC DATALOG

Datalog language used for logical programming in the form [1], [3]

$$\text{head}(X, Y) \text{ :- } \text{body}(X, Y).$$

Manuscript received December 13, 2013; revised March 3, 2014. This work was supported in part by grant from Suranaree University of Technology through the funding of Data Engineering Research Unit.

The authors are with Computer Engineering School Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand (e-mail: mikaiterng@gmail.com).

where  $\text{head}(X, Y)$  is the head of the rule, and  $\text{body}(X, Y)$  is the body of the rule.

Fact is the data is actually stored in the knowledge base and in the form

$$\text{data}(X, Y).$$

Rule is form of fact-based conditions of find the answer. Rule in the form

$$\text{findData}(X, Y) \text{ :- } \text{data}(X, Y), Y < 10,000.$$

## III. DATALOG<sup>TOP-K</sup>

Datalog<sup>top-k</sup> [4], [5] is finding ranking top k in the form

$$q(X, S) \leftarrow \exists y \text{body}(X, Y), S = f(Pi(Zi)) \quad (1)$$

- 1)  $X$  are the distinguished variables;
- 2)  $S$  is the score variable;
- 3)  $Y$  are non-distinguished;
- 4)  $\text{body}(X, Y)$  is disjunctive query of atom;
- 5)  $z_i$  is set of constants or variables;
- 6)  $p_i$  is all fuzzy predicate;
- 7)  $f$  is scoring function taking values in  $[0, 1]$ ;

Fuzzy Set is scope of the fuzzy information and different from crisp, Crisp Set have 2 values of membership degree (true/false or 0/1) but fuzzy set have multiples value and taking values in  $[0, 1]$ , fuzzy set used membership function for indentify value of membership.

Membership function [2], [6], [7] (see Fig. 1- Fig. 4).

### A. Triangular Function

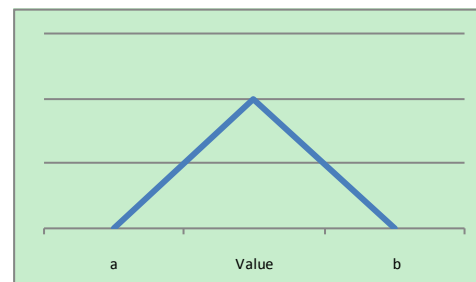


Fig. 1. Triangular function.

$a$  is a lower limit

$b$  is an upper limit

Value is interested value

where  $a < \text{Value} < b$

$$\mu(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{m-a}, & a < x \leq m \\ \frac{b-x}{b-m}, & m < x < b \\ 1, & x \geq b \end{cases} \quad (2)$$

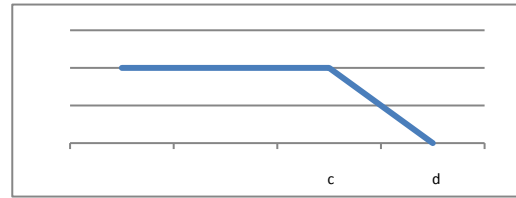


Fig. 4. Right shoulder function.

B. Trapezoidal Function

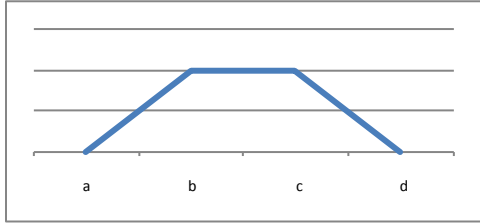


Fig. 2. Trapezoidal Function.

$a$  is a lower limit  
 $d$  is an upper limit  
 $b$  is a lower support limit  
 $c$  is an upper support limit  
 where  $a < b < c < d$

$$\mu(x) = \begin{cases} 0, & (x \leq a) \text{ or } (x > d) \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c < x < d \end{cases} \quad (3)$$

C. Left Shoulder Function

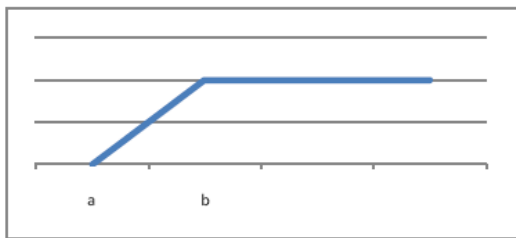


Fig. 3. Left shoulder function.

$a$  is a lower limit  
 $b$  is interested value  
 where  $a < b$

$$\mu(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases} \quad (4)$$

D. Right Shoulder Function

$c$  is interested value  
 $d$  is an upper limit  
 where  $c < d$

$$\mu(x) = \begin{cases} 0, & x < d \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 1, & x > c \end{cases} \quad (5)$$

IV. TOP-K COMPUTING

Datalog<sup>top-k</sup> of population, Researcher selected membership function such as Trapezoidal Function, Triangular Function, Left Shoulder Function and Right Shoulder Function for find ranking top-k answering and select data of population in 2012 from <http://databank.worldbank.org/>, this data have 245 record, in Table I is example data of population 15 record:

TABLE I: EXAMPLE DATA OF POPULATION

Country	Population
Lithuania	2,985,509
Jamaica	2,712,100
Macedonia FYR	2,105,575
Qatar	2,050,514
Kosovo	1,806,366
Armenia	2,969,081
Namibia	2,259,393
Slovenia	2,058,152
Latvia	2,025,473
Gambia	1,791,225
Mongolia	2,796,484
Pacific island small states	2,252,782
Lesotho	2,051,545
Botswana	2,003,910
Guinea-Bissau	1,663,558

A. Triangular Function

TABLE II: POPULATION TOP-10 RANKING WITH TRIANGULAR FUNCTION

Country	Population	Membership Value
Namibia	2,259,393	1
Pacific island small states	2,252,782	1
Macedonia FYR	2,105,575	0.865
Slovenia	2,058,152	0.797
Lesotho	2,051,545	0.788
Qatar	2,050,514	0.786
Latvia	2,025,473	0.751
Botswana	2,003,910	0.720
Jamaica	2,712,100	0.576
Kosovo	1,806,366	0.438

Population top-10 ranking "finding the country which

have the population between 1,500,000 and 3,000,000 people by the population approximate 2,200,000 to 2,500,000 people most" with trapezoidal function in Table II.

Form Table II. After top-k ranking with trapezoidal function. Country which have the population between 1,500,000 and 3,000,000 people by the population approximate 2,200,000 to 2,500,000 people most and have value of membership value is 1 is Namibia which have the population 2,259,393 people and Pacific island small states which have the population 2,252,782 people. Other countries have the value of membership according to the condition.

**B. Trapezoidal Function**

Population top-10 ranking "finding the country which have the population between 1,500,000 and 3,000,000 people by the population approximate to 2,400,000 people most" with triangular function in Table III.

TABLE III: POPULATION TOP-10 RANKING WITH TRAPEZOIDAL FUNCTION

Country	Population	Membership Value
Namibia	2,259,393	0.844
Pacific island small states	2,252,782	0.836
Macedonia FYR	2,105,575	0.673
Slovenia	2,058,152	0.620
Lesotho	2,051,545	0.613
Qatar	2,050,514	0.612
Latvia	2,025,473	0.584
Botswana	2,003,910	0.560
Jamaica	2,712,100	0.480
Kosovo	1,806,366	0.340

Form Table III. After top-k ranking with trapezoidal function. Country which have the population between 1,500,000 and 3,000,000 people by the population approximate to 2,400,000 people most is Namibia, Pacific island small states, Macedonia FYR, Slovenia, Lesotho, Qatar, Latvia, Botswana, Jamaica and Kosovo respectively and have the value of membership according to the condition.

**C. Left Shoulder Function**

TABLE IV: POPULATION TOP-10 RANKING WITH LEFT SHOULDER FUNCTION

Country	Population	Membership Value
Namibia	2,259,393	0.844
Pacific island small states	2,252,782	0.836
Macedonia FYR	2,105,575	0.673
Slovenia	2,058,152	0.620
Lesotho	2,051,545	0.613
Qatar	2,050,514	0.612
Latvia	2,025,473	0.584
Botswana	2,003,910	0.560
Kosovo	1,806,366	0.340
Gambia	1,791,225	0.324

Population top-10 ranking "finding the country which have the population between 1,500,000 and 3,000,000 people by the population approximate to 2,400,000 people most" with left shoulder function in Table IV.

Form Table IV. After top-k ranking with left shoulder function. Country which have the population approximate 2,400,000 people most is Namibia, Pacific island small states, Macedonia FYR, Slovenia, Lesotho, Qatar, Latvia, Botswana, Kosovo and Gambia respectively and have the value of membership according to the condition.

**D. Right Shoulder Function**

Population top-10 ranking "finding the country which have the population between 1,500,000 and 3,000,000 people by the population approximate to 2,400,000 people most" with left shoulder function in Table V.

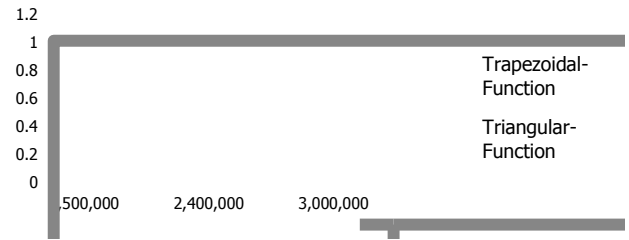


Fig. 1. Difference between trapezoidal function and triangular function.

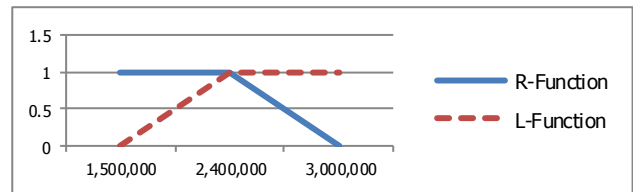


Fig. 2. Difference between trapezoidal function and triangular function.

TABLE V: POPULATION TOP-10 RANKING WITH RIGHT SHOULDER FUNCTION

Country	Population	Membership Value
Jamaica	2,712,100	0.480
Mongolia	2,796,484	0.339
Armenia	2,969,081	0.052
Lithuania	2,985,509	0.024

Form Table V. After top-k ranking with right shoulder function. Country which have the population approximate 2,400,000 people most is Jamaica, Mongolia, Armenia and Lithuania respectively and have the value of membership according to the condition.

**V. CONCLUSION**

Top-k ranking of population. Condition step is very important because condition for find item is a create equation for evaluate membership function. Which membership function is a ranking top-k of population if wrong evaluate of membership value then wrong top-k ranking.

According to experiments can be seen that membership function Trapezoidal Function and Triangular Function will generate in different membership value while both of function have same value direction. The ambiguity information is suitably to sort in condition range. If interested value have one value. Triangular function has accuracy more than Trapezoidal function. If interested value have more one value. Trapezoidal function has accuracy more than

Triangular function (In Fig. 1).

Additionally, membership function Left Shoulder Function and Right Shoulder Function are properly to classify information that less or greater value, respectively (In Fig. 2).

#### REFERENCES

- [1] U. Straccia, "Towards top-k query answering in deductive databases," in *Proc. the 2006 IEEE International Conference on Systems, Man and Cybernetics (SMC-06)*, 2006, pp. 4873–4879.
- [2] J. Dombi, "Membership function as an evaluation," *Fuzzy Sets and Systems*, 1990.
- [3] N. Kerdprasop. (2013). Data mining with datalog. [Online]. Available: <https://sites.google.com/site/nittayak/home/clp>
- [4] U. Straccia, "Towards top-k query answering in description logics: the case of DL-Lite," in *Proc. the 10th European Conference on Logics in Artificial Intelligence (JELIA-06)*, Liverpool, UK, 2006, pp. 439–451.
- [5] U. Straccia, "Answering vague queries in fuzzy DL-Lite," in *Proc. the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-06)*, Paris, 2006, pp. 2238–2245.
- [6] F. Bobillo and U. Straccia, "Fuzzydl: an expressive fuzzy description logic reasoner", in *Proc. the IEEE International Conference on Fuzzy System*, Hong Kong, 2008, pp. 923-930.
- [7] U. Straccia and N. Madrid, "A top-k query answering procedure for fuzzy logic programming," *Fuzzy Sets and Systems*, vol. 205, pp. 1-29, 2012.



**Keerachart Suksut** is currently a master student at the School of Computer Engineering, Suranaree University of Technology, Thailand. He received his bachelor degree in computer engineering from Suranaree University of Technology in 2011. His current research includes top-k ranking with datalog.



**Pasapitch Chujai** is currently a doctoral student at the School of Computer Engineering, Suranaree University of Technology, Thailand. She received her bachelor degree in computer science from Ramkhamhaeng University (RU), Thailand in 2000, and master degree in computer and information technology from King Mongkut's University of Technology Thonburi (KMUTT), Thailand in 2004. Her current research includes ontology, recommendation system.



**Nittaya Kerdprasop** is an associate professor at the School of Computer Engineering, Suranaree University of Technology, Thailand. She received her bachelor degree in radiation techniques from Mahidol University, Thailand in 1985, master degree in computer science from the Prince of Songkla University, Thailand in 1991 and doctoral degree in computer science from Nova Southeastern University, U.S.A in 1999. She is a member of ACM and IEEE Computer Society. Her research of interest includes knowledge discovery in databases, artificial intelligence, logic programming, and intelligent databases.



**Kittisak Kerdprasop** is an associate professor and chair of the School of Computer Engineering, Suranaree University of Technology, Thailand. He received his bachelor degree in mathematics from Srinakarinwirot University, Thailand in 1986, master degree in computer science from the Prince of Songkla University, Thailand in 1991 and doctoral degree in computer science from Nova Southeastern University, U.S.A. in 1999. His current research includes data mining, artificial intelligence, functional and logic programming languages, computational statistics.