Selection of Defuzzification Method to Obtain Crisp Value for Representing Uncertain Data in a Modified Sweep Algorithm

Gunadi W. Nurcahyo
Faculty of Computer Science, University of Putera Indonesia "YPTK"
Padang, West Sumatera 25221, Indonesia
and
Siti Mariyam Shamsuddin, Rose Alinda Alias, Mohd. Noor Md. Sap.
Faculty of Computer Science and Information System, University of Technology Malaysia
Skudai, Johor Bahru 81310, Malaysia

ABSTRACT

We present a study of using fuzzy-based parameters for solving public bus routing problem where demand is uncertain. The fuzzy-based parameters are designed to provide data required by the route selection procedure. The uncertain data are represented as linguistic values which are fully dependent on the user's preference. This paper focuses on the selection of the Defuzzification method to discover the most appropriate method for obtaining crisp values which represent uncertain data. We also present a step by step evaluation showing that the fuzzy-based parameters are capable to represent uncertain data replacing the use of exact data which common route selection algorithms usually use.

Keywords: sweep algorithm, public bus routing, route selection, fuzzy-based parameter, defuzzification.

1. INTRODUCTION

This paper is based on a case study on a public transport company located in Semarang, Indonesia [1]. We propose a solution to public bus routing problem concerning uncertain demand, travel distance, and type of road. The objective is to perform route selection using a modified search technique that is capable to capture uncertain data. A model called Sweep algorithm has been introduced previously for solving route selection problem for goods delivery vehicles [2]. This algorithm performs a route selection using exact data and certain demand. [3] apply the sweep algorithm for solving route selection when demand is uncertain with capacity constraint and additional distance is allowed. In our case, the algorithm is applied to public bus routing problem when uncertain demand is a major concern and no additional distance is allowed. Fuzzy-based parameters are proposed to represent the uncertain data. The fuzzy inference rules are assigned in the fuzzy system to provide data used in the modified sweep algorithm.

This paper focuses on the defuzzification part of the proposed solution where the crisp values are obtained and concerned in the route selection. Defuzzification is a reverse process of Fuzzification. Basically, defuzzification is a mapping from a space of fuzzy control actions defined over an output universe of discourse into a space of non fuzzy (crisp) control actions [4].

2. PROBLEM DEFINITION

A route selection for public bus routing problem is performed concerning demand, travel distance and type of road. Demand is the major concern and in the case of searching for new routes, demand is uncertain. This paper discusses the three parameters required in a route selection process where data are uncertain or, in other words, the data cannot be represented as exact numbers. For known locations of nodes and links in which public bus service does not exist, three conditions of links must be identified approximately and represented as linguistic value such as: very low, low, medium, high, bad, good, etc. The linguistic values are determined based on the user's preference. The conditions of links which become the parameters are namely: *Demand*, *Travel Distance*, and *Type of Road*. Fuzzy inference rules are assigned to convert linguistic values of data into crisp values. The crisp values which represent three parameters are used as the weights of the candidate links. By using the fuzzy-based parameters, the proposed system is designed to capture uncertain data and use them in the route selection process.

3. MODIFIED SWEEP ALGORITHM (MSA)

Modification is made to the second stage of the original sweep algorithm, i.e. Route Generation. In the first stage, clustering process is performed using the same procedure as used in the original sweep algorithm. A node is joined with its nearest neighbour based on its polar coordinate angle and the capacity of each vehicle. The objective of clustering in the modified sweep algorithm is to provide a procedure which enables user to easily group the available nodes and determine which cluster a node or link should be attached to. *Node[0]* is assumed to be the centre polar coordinate.

Since *node[0]* of the real data is located in coordinate $x=324$ and $y=124$, we reformulate the calculation of polar coordinate angles. If the location of *node[0]* is denoted by x_0y_0 then the location of *node[i]* is defined as x_i-x_0,y_i-y_0. The polar coordinate angle is defined using *sinus* function to suit the functions provided by Pascal programming language that is used in the experiment:

$$Sin\ \zeta_i=(y_i-y_0)/radian \quad (1)$$

where radian is defined by:

$$radian=\cdot\ ((x_i-x_0)^2+(y_i-y_0)^2) \quad (2)$$

After clustering process is completed, we have to analyse the list of links in each cluster. As mentioned previously, clustering process enables us to group the links and perform the route selection.

In the second stage, a thorough modification is made by replacing travelling salesman (TSP) route selection method with a weighted-directed search (WIDI). Unlike TSP which selects routes by combining nodes based on short distance, WIDI selects routes by combining links based on demand as explained in Section 5. Comparisons between SA and MSA are presented in Section 7.

In selecting a link to be attached to the generated route, there are several constraints that are considered. Assuming that *l[i,j]* is a link connecting node *i* to node *j* and *L* is a set of links from start node *node[1]* to end node *node[n]*, link *l[i,j]*$\subset$*L* if:

a. *weight[i,j]>=weight[i,j+1]*
   where: 0ΩiΩMaxNode
   0ΩjΩMaxNode
b. *node[i]* and *node[j]* are predecessors of *node[n]*
c. *i*∏*node[n]* and *j*∏*node[n-1]*

The WIDI Graph Search algorithm is described as follows:

1. Initialize *Node*[1] = *Start_Node*, Destination = *End_Node*, Maximum node = *Max_Node*, *m* = maximum adjacent nodes, and *x*$\subset${*Node*[1], *Node*[2], … *Node*[n]}.
2. Store *Node*[1] in the first list of nodes.
3. For *i*=1 to *Max_Node* do
4. *x*=Node[*i*].
5. For *j*=1 to *m* do
6. Find an adjacent node to *x* in the weight matrix.
7. If found and the adjacent node is not in the list of nodes and not in the Adjacent List then store the adjacent node in the Adjacent List.
8. Next *j*
9. Next *i*
10. Initialize *y*=Destination, *t*=false
11. While *y*∏*Start_Node* do
12. for *i*=1 to *Max_Node* do
13. for *j*=1 to *m* do
14. if Node[*i,j*]=y then store Node[*i*] and Node[*i,j*] in the track, *t*=true.
15. next *j*
16. if *t*=true then set the next destination *y*=Node[i]
17. next *i*
18. End while
19. Set an empty stack to store generated links.
20. Start route searching from the *Start_Node* to *End_Node*. Initialize *s*=*Start_Node* and push to the stack.
21. If *s* has more then one adjacent nodes then check whether those adjacent nodes are in the same track or not. If not go to step 23.
22. If yes then compare the weights of those nodes and select the node with highest weight.
23. Push the node to the stack.
24. Set *s* to the selected node.
25. Repeat step 21 until *s*=*End_Node*.
26. The sequence of nodes in the stack represents the links of the selected route and this is the result of the route searching process. To search for another route, repeat step 10.

4. FUZZY PARAMETERS

It is perceived that a candidate route is selected based on three parameters. These parameters coincide with the factors that are used by the company in route selection. Below are the parameters of the candidate routes. Note that each parameter has a range of value from zero (0) to one (1).

a. Demand (D)

Demand is the most concerned parameter of the company in selecting a route. In this research, *demand* is combined with the other attributes as linguistic data for the route selection process. There are 5 membership functions for *demand* that show the degree of potential demand: VERY HIGH, HIGH, MEDIUM, LOW, and VERY LOW denoted as VH, H, M, L, and VL.

Demand at each node is only approximately known and represented as a triangular fuzzy number because this membership representation shows minimum and maximum boundary clearly. Figure 1 shows the membership function of triangular fuzzy number ***D*** representing demand at the link. Triangular fuzzy number ***D*** = (d_1, d_2, d_3) is described by its left boundary d_1 and its right boundary d_3. Thus, the decision maker can estimate that demand at the node will not less than d_1 or greater than d_3 based on his experience and intuition or available data. Figure 1 shows the fuzzy sets of linguistic data for demand.

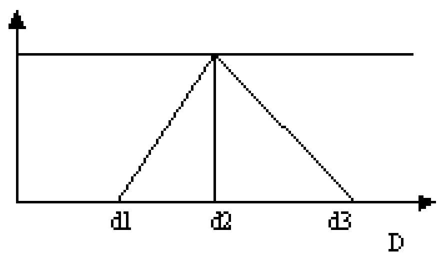


Fig. 1. Triangular Fuzzy Number *D*

It can be seen in Figure 1 that each value has left and right boundary. For example, right boundary of VL is

50. The grade of membership of VL at this point is zero. This means that when demand is 50 or greater, the appropriate linguistic data for demand is Low. At the same point, the grade of membership of L is 1. Hence, if demand is 50 then the linguistic value of demand is LOW. Figure 2 shows the complete membership functions of *demand*.

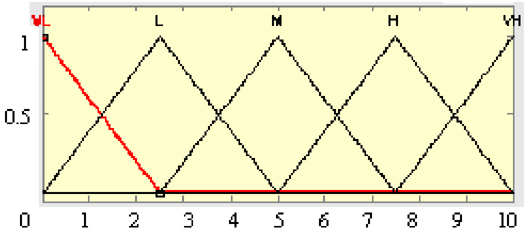


Fig. 2. Demand Membership Functions

b. *Travel Distance (TD)*

Distance of a route is concerned because it is related operational cost. A shorter route requires less operational cost. In public bus routing, a short distance route with minimum cost cannot be selected when demand is low. Therefore, travel distance cannot be used as single parameter for public bus route selection. The linguistic data of travel distance consists of VERY_LONG_DISTANCE (VL), LONG_DISTANCE (LL), MEDIUM_DISTANCE (MD), SHORT_DISTANCE (SS), and VERY_SHORT_DISTANCE (VS). The membership functions of this parameter are shown in Figure 3. The triangular fuzzy number is noted by *TD=(td1, td2, td3)*.

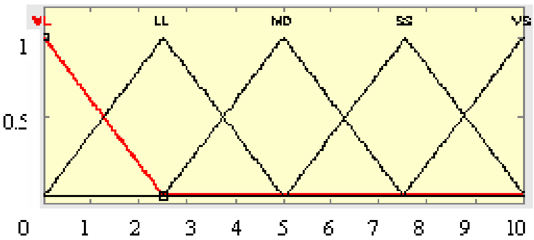


Fig. 3. Distance Membership Functions

c. *Type of Road (TR)*

Type of road includes the physical condition of the roads in a route. A road with good condition enables the buses to minimize travel time. Conversely, a bad road may increase travel time and, in sequence, increases operational cost. Type of road consists of three conditions of road denoted by BAD, AVERAGE, and GOOD with membership functions as shown in Figure 4. Notation for triangular fuzzy number is *TR=(tr1, tr2, tr3)*.

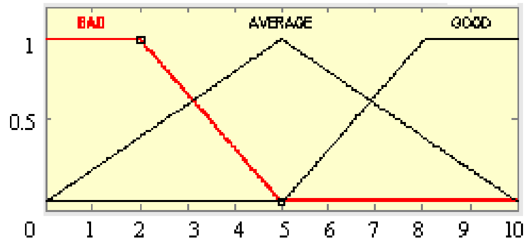


Fig. 4. Type of Road Membership Functions

5. FUZZY RELATION

In this section, we illustrate the relation between the initiated fuzzy input and the required output. Figure 5 shows the whole structure of the proposed solution consisting of input, reasoning rules, and output. Three inputs are captured consisting of *demand*, *distance*, and *type of road* as described previously.

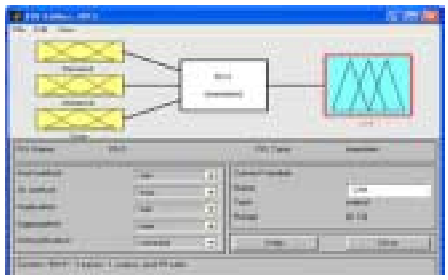


Fig. 5. Fuzzy System Structure

The three inputs are denoted as:

$$\sigma_A(D)$$

where $A = \{VL,L,M,H,VH\}$ (3)

$$\sigma_B(TD)$$

where $B = \{VL,LL,MD,SS,VS\}$ (4)

$$\sigma_C(TR)$$

where $C = \{BAD,AVERAGE,GOOD\}$ (5)

The inference rules relate these input to the output *Link* consisting of seven membership functions i.e. NEGATIVE LARGE (NE), NEGATIVE MEDIUM (NM), NEGATIVE SMALL (NS), ZERO (ZE), POSITIVE SMALL (PS), POSITIVE MEDIUM (PM), and POSITIVE LARGE (PL). In this paper, we use Mamdani type Fuzzy Inference Rules. As noted by [5], every rule represents a fuzzy relation. In our case, every rule represents relation between the three inputs to *Link*. Figure 6 shows the complete membership functions of *Link* which is denoted as:

$$\sigma_O(Link)$$

where $O = \{NL,NM,NS,ZE,PS,PM,PL\}$ (6)

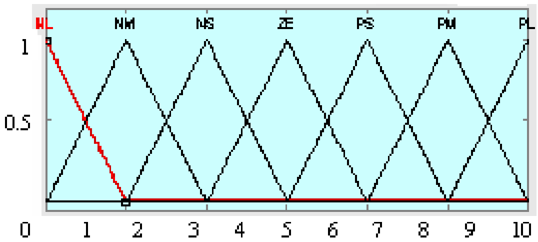


Fig. 6. Link Membership Function

Let us denote by *R* the fuzzy relation between *D*, *TD*, *TR*, and *Link*. As denoted in Teodorovic and Pavkoviv (1996), membership function $\sigma_R(x,y)$ of this fuzzy relation equals:

$$\sigma_R(x,y)=min\{\sigma_A(x), \sigma_B(y)\} \quad \&x,y \qquad (7)$$

In this paper, we denote the fuzzy relation as:

$$\sigma_R(x,y)=min\{min(\sigma_A(D), \sigma_B(TD), \sigma_C(TR)), \sigma_O(Link)\} \quad \&x,y \qquad (8)$$

where *x* is input and *y* is output.

The following rules show the fuzzy relation and are graphically shown in Figure 7:
If [Demand is M] and [Distance is SS] and [Type is BAD] Then [Link is NM]
If [Demand is M] and [Distance is SS] and [Type is AVERAGE] Then [Link is ZE]
If [Demand is M] and [Distance is SS] and [Type is GOOD] Then [Link is ZE]

The fuzzy relations of those rules are shown as follow:

$\sigma_R(x,y)=min\{min(\sigma_M(D), \sigma_{SS}(TD), \sigma_{BAD}(TR)), \sigma_{NM}(Link)\}$

$\sigma_R(x,y)=min\{min(\sigma_M(D), \sigma_{SS}(TD), \sigma_{AVERAGE}(TR)), \sigma_{ZE}(Link)\}$

$\sigma_R(x,y)=min\{min(\sigma_M(D), \sigma_{SS}(TD), \sigma_{GOOD}(TR)), \sigma_{ZE}(Link)\}$

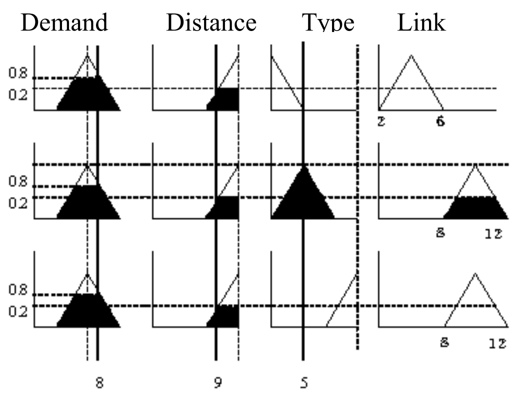


Fig. 7. Fuzzy Relations

6. DEFUZZIFICATION METHODS

The most common defuzzification methods are presented and compared in this paper i.e Center of Area and Maxima methods.

Center of Area/Gravity

It is one of the most commonly used defuzzification techniques. This method determines the centre of the area of the combined membership functions [4]. [6] calculated the centroid or centre of gravity (COG) of the area under the membership function as:

$$F^{-1}_{COG}(\tilde{A}) := \frac{\int_x \sigma \tilde{A}(x) \cdot x \, dx}{\int_x \sigma \tilde{A}(x) \, dx} \qquad (9)$$

Maxima Methods

COG is a defuzzification method regarding the area under the membership function. Maxima methods consider values with maximum membership [6]. There are different maxima methods with different conflict resolution strategies for multiple maxima, e.g., *first of maxima (FOM), last of maxima LOM), mean of maxima (MOM)*, and *centre of maxima (median)*. In this paper, We use MOM and LOM as the representation of maxima methods which are denoted as follow.

Mean of Maxima:

$$U = \sum_{i=1}^{R} u_i / R \qquad (10)$$

Last of Maxima:

$$U = \sup(u\square), \sigma(u\square) = \sup_{u=\subset[Min,Max]} \sigma(u) \qquad (11)$$

7. COMBINING THE SWEEP ALGORITHM WITH FUZZY SYSTEM

Figure 8 illustrates the combination of fuzzy system and the MSA. It can be seen that the combination configures two different levels i.e. linguistic level and technical level. Linguistic level is a level where unknown data accpted from the technical level are converted to the fuzzy numbers. The inference rules relate the three input parameters to the output parameter obtaining the linguistic value of link that will be defuzzified to get the crisp value.

Technical level is a level where a user determines the unknown data and identifies all nodes and their locations to perform the route selection process. The unknown data are brought to the linguistic level for the fuzzification process and the certain data consisting of nodes and locations are required by the MSA for clustering process. Figure 8 shows that there are two types of data required by the route selection i.e. certain data of nodes and locations and the crisp values obtained by the defuzzification process in the linguistic level.

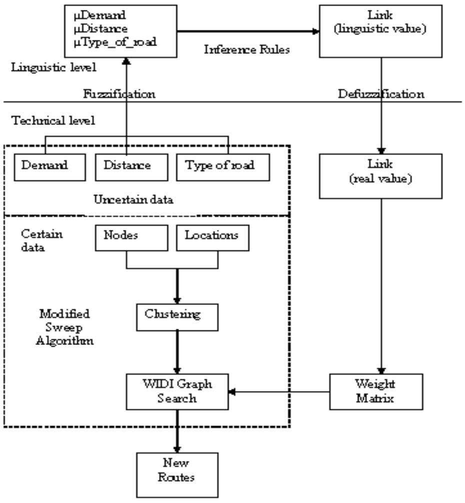


Fig. 8. Combination of Fuzzy System and MSA

8. EXPERIMENTAL RESULTS

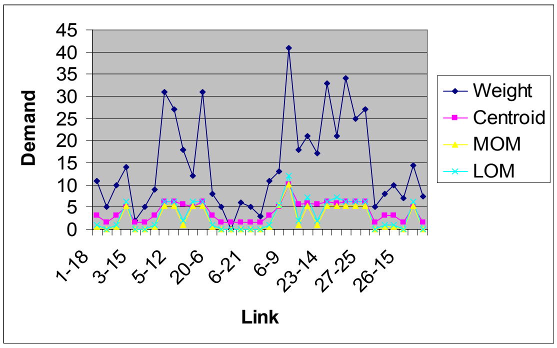
Route selections are performed using both exact and fuzzy using the Modified Sweep Algorithm. There are 27 nodes included in the network and this experiment uses two different presentation of data i.e. numeric and linguistic presentations of the same road network. The selection task aims to generate the best routes which offer maximum demand. Weights of links which represent *demand*, *distance*, and *type of road* are considered in the route selection. Mamdani type Fuzzy Inference System is used to relate input parameters to the output parameter. This section also evaluates three defuzzification methods i.e. Centroid, Mean of Maxima (MOM) and Last of Maxima (LOM) in performing the conversion of linguistic values into crisp values. The three defuzzification methods are evaluated to select the best method to be applied in the algorithm.

The experiment is conducted to investigate the feasibility of fuzzy system in providing data for route selection process. The experiment aims to find evidences that linguistic values representing uncertain data are capable for replacing exact data in public bus routing problem. We evaluate the parameters step by step starting from the most concern parameter i.e. *demand*. A route is selected based on the number of demand in each link. The second-concerned parameter is *distance*. When two candidate links have approximately the same demand, distance parameter is evaluated to identify a link with shorter distance. Note that a candidate link with longer distance may be selected if it has higher number of demand. Type of road is concerned because this parameter is related to operational cost and travel time. Type of road is physical condition of road. A bad condition of road may increase operational cost and travel time. We evaluate the parameters one by one to study their behaviour.

One Parameter

The results of this experiment are compared with the previous results of exact demand data as shown in Figure 9 for alternative links. It is shown in Figure 9a that the fuzzy data pattern may follow the exact data pattern (the weight line). Figure 9b shows only the fuzzy data pattern consisting Centroid, MOM and LOM defuzzification methods.

a)



b)

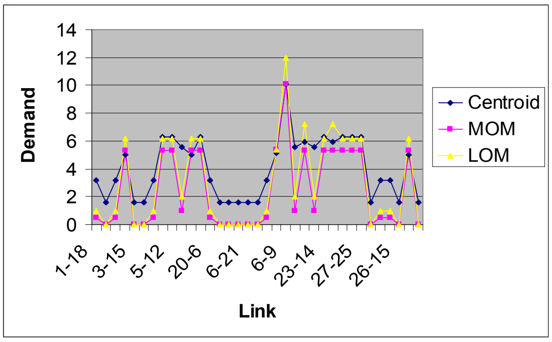


Fig. 9. Demand Patterns

Two Parameters

Based on the experiment using one parameter, it is concluded that fuzzy data is capable to replace exact data indicated by the similar data pattern as shown in Figure 9. In this section, we present the next experiment based on the same case using additional parameter *Distance*. The use of parameter *Demand* and *Distance* in this experiment aim to obtain more accurate results which represent the real condition of public bus routing problem. The objective of this experiment is to evaluate the capability of fuzzy system in providing weight of link required by the MSA. The weight data represents the linguistic values of both *Demand* and *Distance*. The fuzzy relations between *Demand*, *Distance*, and *Link* are presented in Fuzzy Associative Matrix (FAM) as shown in Figure 10. FAM shows 25 rules relating the fuzzy input to the output. A sample of the results is shown graphically in Figure 11.

Travel Distance

Demand	VL	LL	MD	SS	VS
VL	NL	NL	NL	NL	NM
L	MN	NM	NM	NS	NS
M	NS	ZE	ZE	ZE	ZE
H	PS	PS	PS	PM	PM
VH	PM	PM	PL	PL	PL

Fig. 10. Fuzzy Associative Matrix

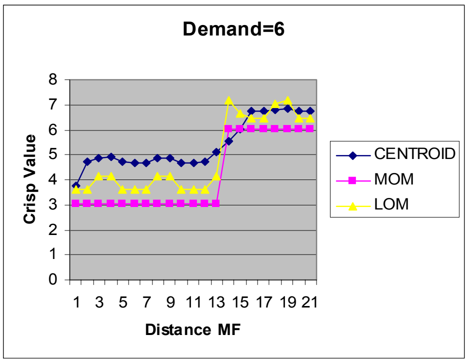


Fig. 11. Data Patterns of Demand = 6

In the first six results, i.e. *demandÒ5*, Centroid , MOM and LOM methods have the same data pattern. At certain point, the graph raises drastically separating the lower parts with the higher parts. When *demandÓ6* the results of MOM and LOM still follow the same pattern. The results of Centroid, on the other hand, move to different pattern where there are more frequent changes of crisp values. This type of data pattern is more representative for providing weight data required by the proposed model.

It is shown in Figure 8 that, in general, the crisp value increases as the value of *Demand* and *Distance* membership functions increase. Note that the value of *Distance* membership functions does not represent the actual distance. As shown in the order of fuzzy set *Link* (see notation 4), the leftmost membership function represents the longest travel distance and vice versa. The *Distance* parameter is required to select a link when two candidate links have approximately the same demand. In our case, short distance is not a major concern. A candidate link may be selected even if it has longer distance than the other candidate. Let us see the following results:

Result 1:

$\sigma_M(D)=10,\ \sigma_{SS}(TD)=15,\ \sigma_{ZE}(Link)=9.01$

Result 2:

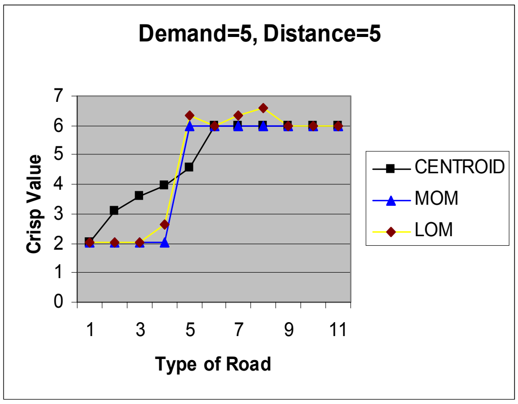
$\sigma_H(D)=15,\ \sigma_{LL}(TD)=5,\ \sigma_{PS}(Link)=12$

Both results represent two candidate links. Result 1 is a link that has medium demand and short distance. Result 2, on the other hand, is a link that has high demand and long distance. In this case, although it has longer distance than result 1, result 2 is selected because it has higher demand.

Three Parameters

The results of this experiment share the same pattern as those of the previous experiment. The Centroid method also produces more representative results than MOM and LOM because there is a gradual increase of crisp values for certain membership function of *Demand* and *Distance*. Table 2 lists several instances of fuzzy results using three parameters. By concerning parameter *Type of Road*, the MSA is capable of avoiding bad links even if those links have higher number of demand. Let us compare case 14 and case 22 for example. Although the demand of case 22 is very high, case 14 is better because it has a better condition of road even if the demand is medium. A better link is indicated by the higher crisp value of parameter *Link*. Figure 11 shows the two cases in graphical forms.

a)



b)

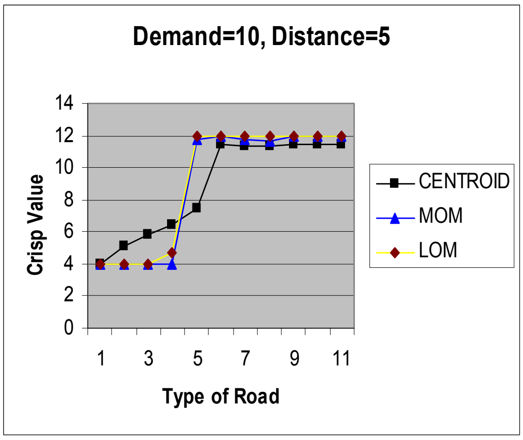


Figure 11: Data Patterns of Demand = 6

Table 2: Fuzzy Results using Three Parameters

CASE	DEMAND	DISTANCE	TYPE	CRISP VALUE
1	Very Low	Very Long	Bad	0.632
2			Average	0.632
3			Good	0.632
4		Medium	Bad	0.632
5			Average	0.632
6			Good	0.632
7		Very Short	Bad	0.632
8			Average	2.02
9			Good	2.02
10	Medium	Very Long	Bad	0.632
11			Average	4.01
12			Good	4.01
13		Medium	Bad	2.02
14			Average	6.01
15			Good	6.01
16		Very Short	Bad	2.02
17			Average	6.01
18			Good	6.01
19	Very High	Very Long	Bad	4.01
20			Average	10
21			Good	10
22		Medium	Bad	4.01
23			Average	11.4
24			Good	11.4
25		Very Short	Bad	4.01
26			Average	11.4
27			Good	11.4

9. CONCLUSIONS

Centroid method is chosen to perform the defuzzification process based on the comparison of three well-known defuzzification methods i.e. Centroid, MOM and LOM methods. Centroid defuzzification method converts the Link membership functions into crisp values. The crisp values are assigned as the weights of the candidate links which are selected by the algorithm to generate route. A selected link is indicated by higher crisp value than another candidate. It is also demonstrated that the fuzzy-based parameters are capable of representing uncertain data used by the MSA.

10. REFERENCES

[1] G. W. Nurcahyo, Siti Mariyam Shamsuddin, Rose Alinda Alias, M. Noor Md. Sap, "Vehicle Routing Problem for Public Transport: A Case Study", International Technical Conf. on Computers/Systems, Circuits and Communications, Vol. 2, pp. 1180-1183, Phuket, Thailand, 16-19 July 2002.

[2] B. E. Gillet, L. R. Miller, "A Heuristic Algorithm for the Vehicle Dispatch Problem", Operation Research, Vol. 22, 1974, pp. 340-349.

[3] D. Teodorovic, G. Pavkovic, "The fuzzy set theory approach to the vehicle routing problem when demand at nodes is uncertain", Fuzzy Sets and Systems, Vol. 82, 1996, pp. 307-317.

[4] D. H. Rao, S. S. Saraf, "Study of Defuzzification Methods of Fuzzy Logic Controller for Speed Control of a DC Motor", IEEE Transactions, 1995, pp. 782-787.

[5] E. Mamdani, S. Assilian, *An experiment in linguistic synthesis with a fuzzy logic controller*, Internat. J. Man-Machine Studies, Vol. 7, 1975, pp. 1-13.

[6] T. A. Runkler, "Extended Defuzzification Methods and Their Properties", IEEE Transactions, 1996, pp. 694-700.