Online Decision Support System of Used Car Selection Using K-Nearest Neighbor Technique

Sasitorn Kaewman, Wallapa Khemsanthia, Oratai Boongomud, and Chatklaw Jareanpon

***Abstract*—This paper, we present Decision Support Systems (DSS) for used car selection using k-nearest neighbor (k-NN) algorithm. This system presents a user-friendly web-based spatial decision support system aimed for selling vehicle. There are two functions that are searching using k-nearest neighbor algorithm, and searching in the database to find one or more matching datasets with the user query. For a web search query, records contain vehicle type, brand, model, year, size, prince and type of engine. By contrast, k-nearest neighbor was estimated from data normalized of the training data set. The result showed that (DSS) present highly effective and sustainable tools for searching vehicle.**

***Index Terms*—K-nearest neighbor algorithm, normalized, search query, web application.**

I. Introduction

Before buying a car, user must get some information or study about what kind of car is suitable for them. Basically, the buyer who wants to buy a new car model can visit the showroom to get more information. In used car, some model is very old and out of the showroom. Moreover, nowadays, internet can be used to get the information. The web database is finding one or more matching datasets with the user query. In this system, user can query the records that are able to select the suitable car. By contrast, the Decision support system (DSS) is the method that will make decision according to the information or feed back that are key in by users.

The definition support system is defined by [1], [2], "Decision are unprogrammed to the extent that they are novel, unstructured and unusually consequential. There is no cut-and-dried method for handling the problem because it has no arisen before, or because its precise nature and structure are elusive or complex, or because it is so important that it deserves custom-tailored treatment."

From this definition, many applications are applied by DSS such as:

DSS for Real Estate [3] is build up for area development or building according to the rules. The prospective buyer will choose the house sites or building based on their choice. It will develop to become an information area system for construction. This system will also allow one to choose the dwelling units according to their choice.

[4] is designing and implementing a web-based DSS system called PlanIT online. This system was intended to be comprehensible for set group for person with disabilities and provides information about work incentives andbenefits from social employment policies. Web service will be used to compute between government agencies in federal and state level. The main objective of this project is to integrate the two key concepts that are to create a higher quality of information and to make it easily.

From the problem on selected the used car and the effectiveness of DSS, we proposed and developed the DSS online to select the used car that decision underlying the user and analyze data generated from transaction processing system easily. Moreover, it is very helpful to friendly support the user.

Manuscript received April 10, 2012; revised May 5, 2012.

The authors are with Department of Computer Science, Faculty of Informatics, Mahasarakham University, Thailand.

II. The Online Decision Support System of Used Car Selection Using K-Nearest Neighbor Technique

The online decision support system of used car selection using K-nearest neighbor technique (K-NN) is divided into 3 parts that are data collection, data preparation, and decision making.

A. Data Collection

TABLE I: The Output of the String Normalization Step.

Ind	Type	Brand	Model	CC	Price	Gear	Fuel	Color	year
1	0	1	1	1600	465,000	0	1	1	1998
2	0	1	1	2300	275,000	0	1	1	1995
3	1	1	1	3000	199,000	1	0	1	2004
4	0	0	0	2000	549,000	0	0	1	2006
5	1	1	1	2800	179,000	0	1	1	2001
6	0	1	1	4300	765,000	1	0	1	2003
7	0	1	1	1600	765,000	0	1	1	2003
8	0	0	1	1800	205,000	0	0	0	1999
9	0	1	1	1500	399,000	0	0	1	1997
10	0	1	1	1700	429,000	1	0	1	1998
11	1	0	1	2700	435,000	0	0	1	2006
12	1	1	1	2000	419,000	0	0	0	2005

The data of used car is collected from the used car web in THAILAND such as www.one2car.com, www.unseencar,com. However, the format of each web is different. We collect the 500 recodes of used car, and it divided into 2 groups: 373 of compact cars and 142 of SUV cars. The popular key that user usually search are type, brand, model, year, gear, size of machine, fuel, car detail, price, seller, and picture. From 13 keywords, we selected to used only 9 keywords that are most important: brand, model, year, price, gear, size of machine and fuel to make the model answer of user requirement for K-NN algorithm.

B. Data Preparation

From the different data source, the data must be normalization to same format. In this step, we divided into 2

steps: String normalization and Integer normalization.

C. String Normalization

In this step, we compare the string of input and the data in database. The string normalization is the word from brand, type, model, gear, fuel, color, year, and cost, for example the input data is ("TOYOTA, compact, Camry, Auto, petrol, gold, 2001, (800,000-1,000,000 bath)). The string normalization is calculated from

$$Different(x_i, y_i) = \begin{cases} o, if\ (x_i = y_i) \\ 1, otherwise \end{cases} \quad (1)$$

where x_i is the input data from user and y_i is the data from database

From this normalization step, the output is shown in Table1.

After this step, some attributes is not in form of normalization such as size of machine and year. We will transform that attributes using Integer normalization step.

D. Integer Normalization

For K-NN algorithm, the integer normalization will transform the number to only in range 0-1. The size of machine and year are calculated by (2).

$$Integer\_normalization = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

where x is the data that will normalize, min(x) is the minimum of set x, and max(x) is the maximum of set x. For example, from Table 1, the type of the machine (CC) in the first recode (x equal to 1600). Integer normalization is equal to 0.357. For example, the year in first recode equal to 1998, after normalization step the output is 0.214. The integer normalization is shown is Table II.

TABLE II: THE OUTPUT OF THE INTEGER NORMALIZATION STEP

Ind.	Type	Brand	Model	CC	Price	Gear	Fuel	Color	year
1	0	1	1	0.357	1	0	1	1	0.214
2	0	1	1	0.286	1	0	1	1	0
3	1	1	1	0.536	1	1	0	1	0.643
4	0	0	0	0.178	1	0	0	1	0.786
5	1	1	1	0.464	1	0	1	1	0.428
6	0	1	1	1	1	1	0	1	0.571
7	0	1	1	0.036	1	0	1	1	0.571
8	0	0	1	0.107	1	0	0	0	0.286
9	0	1	1	0	1	0	0	1	0.143
10	0	1	1	0.071	1	1	0	1	0.214
11	1	0	1	0.428	1	0	0	1	0.786
12	1	1	1	0.178	1	0	0	0	0.714

After this step, all attributes data is in normalization form. The next step, data will feed to the K-NN algorithm.

E. K-NN Algorithm

The data will feed to the K-NN algorithm that will calculated from the following step.

1) Setting K that K is the number of the result.
2) Calculating the similarity of the input data and the data using this equation

$$D(x, y) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \quad (3)$$

3) Sorting the output

The pseudo-code is shown in Fig1.

```
START
INPUT FROM   Keyword User
INPUT K
DO WHILE NOT K
                READ used car data FROM DATABASE
    //  find the similarity of input using (3)

D(x, y) = sqrt( sum_{i=1}^{m} (x_i - y_i)^2 )

END DO
SORT K
PRINT Output (suitable used car)  IN DISPLAY
END
```

Fig. 1. Pseudo-code of this proposes.

For example input, (Compact car, TOYOTA, Camry, 1800, (800,000-1,000,000), auto, petrol, Gold, 2001), the output of K-NN algorithm is shown in Table III.

TABLE III: THE OUTPUT OF THE PROPOSED ALGORITHM

Ind.	Type	Brand	Model	CC	Price	Gear	Fuel	Color	year	Result
1	0	0	1	0	1	0	0	0	0.005	1.415
2	0	0	0	0.005	1	0	0	1	0.327	1.527
3	0	1	1	0.011	1	0	0	1	0.005	2.003
12	1	1	1	0.005	1	0	0	0	0.25	2.063
4	1	0	1	0.103	1	0	0	1	0.327	2.105

We will ascend sorting the output. The most suitable output has minimized similarity.

TABLE IV: EXPERIMENT RESULT

Input keyword	K-NN	General search
1. No input	100%	100%
2. Complete input	100%	100%
3. Type + Brand + Model	100%	100%
4. Type + Brand + Model + Size of machine + Fuel + Year	83.33%	0%
5. Brand + Model + Size of machine + Year + Color	83.33%	0%
6. Brand + Model + Size of machine + Year + price	83.33%	0%

III. GENERAL SEARCH

The general search is query the requirement of user from the database. If the result cannot find in the database, the output is not display.

IV. EXPERIMENTAL RESULT AND CONCLUSION

From the previous example, the experiment will show in

Table IV.

From the questionnaire of satisfaction of 100 users that are 56 of men and 44 of women, we scale the score in range 1-5 ordering by satisfaction. The score of most satisfaction (5) is equal to 82.86%, the score of (4) is 10.86%, the score of (3) is 4.85%, the score of (2) is 0.85% and the score of (1) is 0.57%.

From the result of satisfaction and the experimental result, the web application DSS using K-MM is suitable and easy for the requirement of user than the general search.

V. Future Work

The future work is possible to compare with the other algorithm such as Decision tree C5.5.

References

[1] H. A. Simon, "The New Science of Management Decision," Harper Collins publisher, English, January 1960.

[2] J.-C. Pomeral and Frederic Adam, "Understanding the Legacy of Herbert Simon to Decision Support Systems," *IGI Global*, pp. 930-938, 2008.

[3] R. R. Trippi, "Decision Support and Expert Systems for Real Estate Investment Decisions: A Review," *Interfaces* vol. 20, no. 5, pp.50-60, 1990.

[4] K. Beak, "The PlanIT online Project: A Web-Based Work Transition Decision Support System for Persons with Desabilites," An Honours project report for COMP 4905 School of Computer Science Carleton University, Ottawa, 2004.