7th International Conference on Innovation in
SCIENCE AND TECHNOLOGY
23 - 25 October, 2020 Amsterdam, Netherlands

Providing a Model for Designing a Decision-Making System Using Fuzzy Opinion Mining Process

Majid Jangizehi[1]*, Javad Hosseinkhani[2], Abdolreza Rasouli Kenari[3], Asma Roshandel Hormozi[2]

[1]Department of Computer Engineering, Sistan and Baluchestan Science and Research Branch, Islamic Azad University, Zahedan, Iran
[2]Department of Computer Engineering, Zahedan Branch, Islamic Azad University, Zahedan, Iran
[3]Faculty of Electrical and Computer Engineering, Qom University of Technology, Qom, Iran

Abstract

Nowadays, the development of information and communication processes has caused organizations to obtain the views of a wide range of their users by various ways and using methods such as data mining and more specifically, web mining, make this data meaningful ultimately exploring knowledge through this data. This research, as one of the main parts of web mining, deals with the subject of opinion mininig and emotion analysis. The main background of this research is the undeniable ambiguity in the heart of people's beliefs and feelings; because most people express their opinions and ideas using verbal variables that are inherently ambiguous. On the other hand, the existence of such ambiguity makes the managers of the organization hesitate to make timely and appropriate decisions, since they are not aware of the main purpose of users. Therefore, in the present study, a model based on fuzzy theory has been proposed to remove the ambiguity of people's opinions and feelings in order to provide a suitable decision in the organization. Afterwards, with the use of the toolbox of MATLAB software a decision-making issue in which the opinions of individuals expressed with verbal variables are disambiguated. The results show that fuzzy theory is a suitable tool for disambiguating the ideas and feelings and is very effective when making decisions among a wide range of opinions expressed with verbal variables.

Keywords: Data mining, web mining, opinion mininig, fuzzy theory, decision Disambiguating

1. Iintroduction

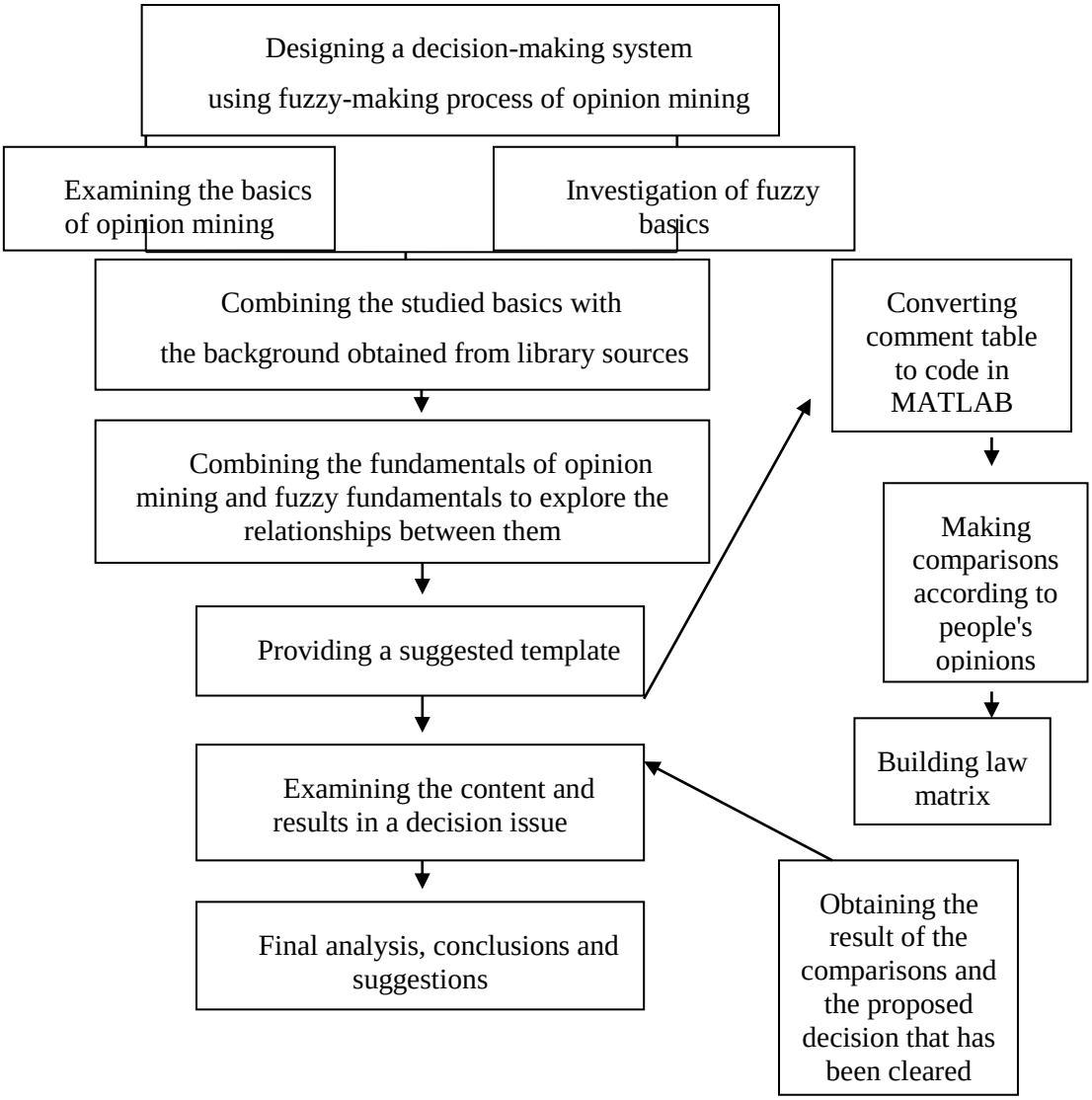
In today's world, with the development of data and information systems, various organizations are looking for an efficient tool to process, analyze, extract knowledge and ultimately infer from the knowledge found. One of the ways in which this can be achieved is through data mining. Data mining emphasizes pattern discovery through large volumes of data with minimal user intervention (Hosseinkhani, 2013). On the other hand, due to the undeniable role of the web, the constant presence of people in an unlimited space and the vast amount of data and information in it, the need for data mining in the web space to extract existing patterns and knowledge becomes necessary. Such an operation is called web mining (Hosseinkhani, 2014). One of the parts of web mining is analyzing people's opinions and

feelings. This is important because by analyzing people's opinions and feelings, one can understand the needs, attitudes and goals of people for being on the web. However, one of the problems in this field is the way of analyzing people's opinions and feelings since people express their speech in natural language and with the help of verbal variables, on the other hand, natural language and verbal variables are ambiguous in nature. Therefore, it is necessary to use a tool to disambiguate people's opinions, ideas and feelings (Bojadziev & Bojadziev, 2002). One of these tools is fuzzy theory (Asqarpour, 2004). Fuzzy view acts in a way so that in a process-oriented process people's ideas and feelings are taken out of the form of verbal and qualitative variables and expressed quantitatively in the form of numbers and even unambiguous verbal variables (Zadeh, 2003). Due to the impact of opinions, ideas and feelings of people on the web on the type and decision-making tendencies of managers of organizations and the lack of sufficient research in the field of analysis of ideas and feelings, the need for such research is felt. In terms of perspective the current research is of theoretical-applied type and is descriptive and survey-based in terms of method of doing (Lesani, 2011).
This research seeks to answer the question of how to dispel ambiguities of ideas and feelings, and whether the fuzzy perspective is a suitable tool for resolving this ambiguity or not?
Therefore, the main objectives of this study are:

1. Exploring the basics of opinion mininig and emotions in the web space
2. Using fuzzy perspective to identify ambiguous aspects of opinion mininig and debugging it
3. More familiarization with users' opinions and thus the realization of the principle of user satisfaction and profitability of the organization
4. Minimize inefficient decisions by members of the organization

1.1. Theoretical Framework of Research

Diagram 1: Theoretical Framework of Research



1.2. Research Background

Various researches have been done in the field of opinion mininig principles, its effect on the mutual relations of the members of the organization with the aim of removing ambiguity from the opinions and feelings of individuals: categorizing and deciding about the opinions of Twitter users and discovering semantic differences and similarities (Madani et al., 2020), designing a decision algorithm using deep thinking-based learning (Yang et al., 2020), using a fuzzy approach to solve the problem of ranking organizations based on effective parameters in e-commerce based on the analysis of users' feelings and opinions (Velvizhy et al., 2020).

In another article, Lotfizadeh examined search engines, specifically Google, for answering searchers' questions, identifying content, understanding requests, and summarizing to meet

users' needs (Zadeh, 2003). Also, Al-Mimani et al. have described the process of opinion mininig from two perspectives: semantics and the possibility of removing ambiguity with fuzzy theory (Al-Maimani et al., 2014). In addition, Jusoh and Alfawareh provided a model for improving the level of analysis of people's beliefs and feelings based on information extraction, knowledge discovery in a fuzzy environment (Jusoh & Alfawareh, 2013) and Hong et al. proposed an integrated model based on fuzzy genetics to increase the efficiency of the opinion mininig process (Hong et al., 2014). Also Casillas et al. proposed a model for extracting ambiguous and questionable data on consumer behavior using a multi-functional system based on fuzzy genetics (Casillas & Martínez, 2009).

1.3. Fundamentals of Opinion Mininig and Investigating Emotions

1.3.1. Data Mining

Data mining has been called the science of discovering knowledge from a database. The result of data mining is the discovery of useful patterns or knowledge from various information sources (Hosseinkhani, 2013).
Data mining is done in three main steps:

1. Pre-processing: At this stage, many unsuitable data for data mining are removed.
2. Data mining: In this step, on the selected data from the previous step, data mining operations are performed and a suitable pattern is generated or discovered.
3. Post-processing: Some of the patterns produced in the previous step are not useful or at least do not have the necessary efficiency for the specific application that is intended. In this step, the useless patterns are eliminated.

The three operations are performed repeatedly and are eventually generated and discovered from raw data, pattern or useful patterns (Pang & Lee, 2008).

1.3.2. Web Mining

Web mining is the discovery of useful patterns and knowledge intended in data mining; The difference is that such a goal is achieved in data mining in the midst of large volumes of data and in web mining in the structure of links and web content (Hosseinkhani, 2014). Web mining consists of three main parts: structure mining, content mining and application mining. Mining structure extracts useful knowledge from the structure of web links; That is why the structure web mining is one of the main technologies in the organization of any search engine. At this stage, only the structures and layout of web pages are examined and discovered (Al-Maimani et al., 2014).
Content Mining: This step extracts and explores useful and tailored knowledge based on needs from the content of web pages. This step helps search engines to automatically categorize web pages based on parameters such as subject, content, and so on. Mining application detects user access patterns by recording clicks. In other words, mining application tries to infer information about user needs by examining user interactions and web page (Hosseinkhani, 2014). Content mining itself is examined from three aspects, including extraction of structured data, information integrity, and opinion mininig. Extraction of structured data examines important web page information, such as a list of services to get feedback from users and provide various services such as value-added and comparative

purchases. Information integration investigates information received from different web pages in order to create a consistent and coherent database to provide services to users (Cambria et al., 2013).
Opinion mining, as the third aspect of content mining, is discussed in the next section.

1.3.3. Opinion Mining

Opinion mining is a way to retrieve and extract knowledge from the context by the use of data mining and processing natural language (Seerat & Azam, 2012). The main function of opinion mining is investigating unstructured text on the web and analyzing beliefs and emotions of the users of the web. One of the most significant objectives of opinion mining is the recognition and expression of people's ideas and emotions by the use of computer. This is why opinion mining is also called a tool for analysis of emotions. Every year large organizations spend huge amounts of money to analyze the feelings and opinions of their customers in order to identify their needs and motives. Gathering the opinions of people through polling centers, etc. is done easily. However, the main point of analyzing these ideas is to discover and extract knowledge and eventually apply the extracted knowledge in organizations. On the other hand, extracting knowledge from web pages is a challenging task, because the data on the web is dynamic and is constantly changing due to constant updates, additions and deletions. Consequently, various organizations are interested in obtaining, albeit heavy costs, the opinions of different people to make decisions in the internal and external areas of the organization, and achieve knowledge that can be processed and generalized by machine. (Zadeh, 1999).
In most sources, opinion mining and emotion analysis are used equivalently; nonetheless, it should be noted that opinion mininig emphasizes the recognition of tendencies and analysis and perception of emotions. Both utilize data mining and natural language processing to discover, retrieve, and use information and ideas on the web (Dalal & Zaveri, 2014).

1.4. Fuzzy Basics

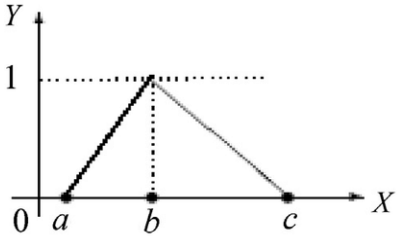
1.4.1. Fuzzy Logic

Fuzzy logic is a useful tool for controlling complex subsystems and processes, detection systems, expert systems, and more. This logic was presented in 1965 by Professor Lotfizadeh. The introduction of fuzzy logic solved the shortcomings of Boolean two-state logic; because many problems in the real world cannot be described and solved in a two-dimensional state and need to be examined in the intermediate state (Nayak & Dash, 2011).
In fuzzy mode, if X is a desired reference set, the fuzzy subset A of the set X gives each member, such as 'x', a value ranging from zero to one. This function is represented by A(x)μ and is called the membership function A or membership degree μ in A (Rajabzade, 2014).
Each member of the desired set A can be made fuzzy with one of the types of conventional sets. The simplest form of representation is a triangular fuzzy number displayed with the membership function A(x)μ on R, along with three real numbers as F= (a, b, c). The upper bound, denoted by 'c', is the maximum value that the fuzzy number F can take. The lower bound denoted by 'a' is the minimum values that the fuzzy number F can take. The amount of

'b' is the most probable value of a fuzzy number (Kahraman, 2015). The membership and image function of a triangular fuzzy number is as follows in Fig 1:

Figure 1, Triangular fuzzy number membership function

$$\mu_A(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a < x \le b \\ 1, & x = b \\ \dfrac{c-x}{c-b}, & b < x \le c \\ 0, & x \ge c. \end{cases}$$



(Zadeh, 1997)

When fuzzy operations are performed on values, the outputs are fuzzy values. These fuzzy results are not easy to understand and interpret, so they must be converted to definite (ordinary) numbers in a process. This process is called de-fuzzing. There are several ways to convert fuzzy numbers to definite numbers. One of these is the marked distance method. This method is used to de-fuzzy triangular fuzzy numbers using equation (1):

$$D(A) = (a + 2b + c) /4 \qquad (1)$$

Other methods such as center of gravity are also used to de-fuzzy (Ariyanejad & Safakish, 2009).

1.4.2. Verbal Variables

The different states that a language variable can take on are called linguistic conditions, or conditions for short. Linguistic conditions are linguistic and schematic interpretations of different situations that a parameter with an event can technically take (Nayak & Dash, 2011). In real world, it is difficult and vague to understand and deduce verbal variables. In such cases, it is better to use numerical variables appropriate to each verbal variable to facilitate the problem solving process. In most cases, the hourly scale is used to equate verbal variables. Table (1) shows the types of verbal variables and their numerical equivalents on an hourly scale along with their fuzzy form:

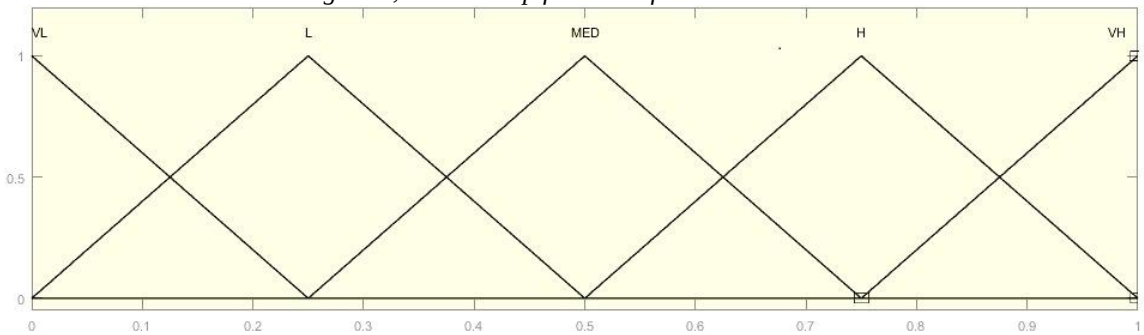
Table 1, Verbal variables on an Saaty scale

Types of preferences (Verbal variables)	The numerical equivalent of crisp	Equivalent of a triangular fuzzy number
Too much preference , very important, very desirable, very much	5	(0.75,1,1)
High preference, important, desirable, high	4	(0.5,0.75,1)
Medium preference, moderately important, moderately desirable	3	(0.25.0.5,0.75)
Low preference, low importance or low desirability	2	(0,0.25,0.5)
Very little preference, very little importance, very little desirability	1	(0,0,0.25)

(Bojadziev & Bojadziev, 2002)

Figure (3) shows a triangular membership function of these verbal variables:

Figure 2, membership function of verbal variables



(Jangizehi et al., 2018)

2. Introducing the Proposed Model and Results

In the previous sections it was said that most people are not explicit in expressing their opinions and feelings and use verbal variables that are ambiguous in nature; thus the process of opinion mininig and emotion analysis was introduced as an inherently ambiguous process and fuzzy theory was introduced as an effective tool to resolve such ambiguities.

Fuzzy theory, by affecting three main criteria, removes the ambiguity of opinion mininig:

1. Emotions: Fuzzying emotions that are expressed with verbal variables, performing the necessary operations and finally de-fuzzying them, removes a degree of ambiguity of emotions (Sudhakaran et al., 2013).
2. Manipulation of text features: Fuzzy theory by manipulating and displaying text features, determines the author's main purpose of expressing the text. It can be said that this process is used to analyze the author's opinions and feelings (Martens et al., 2008).
3. Comparing and exploring the relationships between sentences: fuzzy theory is suitable for analyzing cause and effect relationships; therefore, the rate of text detection increases if fuzzy theory is used (Zadeh, 1997).

2.1. Suggested Model

In this section, a proposed model for disambiguating the process of opinion mininig and emotion analysis is peoposed using fuzzy theory:

1. Collecting ideas, sentences and feelings or whatever the person expresses: At this stage, the person is free to express their opinion in the desired tone.
2. Obtaining and extracting parts of the text that are expressed with more emphasis using verbal variables.
3. Converting verbal variables to their numerical equivalents using methods such as triangular fuzzy numbers.
4. Operation on a set of triangular fuzzy numbers obtained, if necessary, especially when one has used the if-then rules to express one's opinions

5. Finalizing the obtained fuzzy set or sets: This section is done to unify the obtained sets.
6. De-fuzzying the obtained fuzzy sets: In this section opinions, ideas and feelings are displayed as normal variables and previous crisp, with the difference that their ambiguity has been reduced.

This model is introduced based on a review of a number of sources and a combination of the results of previous articles (Al-Maimani et al., 2014), (Hosseinkhani, 2014), (Jusoh & Alfawareh, 2013), (Du & Tan, 2009), (Su et al., 2008), (Bojadziev & Bojadziev, 2002).

2.1.1. Investigation of a Practical Example with MATLAB Software

In this issue, users' opinions about Mahan Airlines (IRM), Caspian (CPN), Homa (IR) and Taban (TBN) based on the five criteria of ticket price appropriateness (TP), fleet quality (FQ), customer orientation (CO), quality of reception (CQ) and number of ticket offices (TO) were received. These criteria are based on user feedback and review of previous research (Jangizehi et al., 2018), (Asfe et al., 2014), (Jangizehi et al., 2013). The goal is to find out the name of the airline with the most acceptance among users, which is expressed with ambiguous verbal variables. For each of these criteria, with the help of user feedback, a weighting factor is considered. Also, for each of the criteria and the names of the airlines, based on the opinions of the users, in pairs, a degree of impact is mentioned. Weight and impact rate are expressed as verbal variables namely very high (VH), high (H), medium (M), low (L) and very low (VL).

The criteria and option for solving the problem, along with the degrees of impact of each are given in Table (2):

Table 2, introduction of the hypothetical problem

Weight factor	H	VH	MED	VH	MED
Criteria / Airline name	TP	FQ	CO	CQ	TO
IRM	L	H	H	L	VL
CPN	VL	L	VH	VL	L
IR	MED	VH	VH	H	VH
TBN	MED	MED	H	VH	MED

Data's Problem and assumptions

After determining the opinions of users based on the criteria of the problem and rates of impact, the problem is defined in MATLAB software. In defining the verbal variables of the problem, the hourly scale ranging from zero to one has been used. The codes written in MATLAB software including the definition of the global set, the verbal variables as well as their membership degree and the number obtained for each airline are given in Appendix One.

2.1.2. The Result of Implementing the Proposed Model

The result of implementing the proposed model in the form of fuzzy decision based on the conditions stated is as follows:

u=
0.8800 0.8000 0.7200 0.6400 0.5600 0.4800 0.4000 0.3200 0.2400 0.1600 0.0800 0
0.9600
vl=
0.6711 0.7223 0.7742 0.8254 0.8738 0.9174 0.9538 0.9808 0.9964 0.9996 0.9901
0.5748 0.6217
l=
0.7353 0.7872 0.8378 0.8853 0.9273 0.9615 0.9858 0.9984 0.9984 0.9858 0.9615
0.6339 0.6838
h=
0.9615 0.9858 0.9984 0.9984 0.9858 0.9615 0.9273 0.8853 0.8378 0.7872 0.7353
0.8853 0.9273
med=
0.8621 0.9071 0.9455 0.9750 0.9936 1.0000 0.9936 0.9750 0.9455 0.9071 0.8621
0.7613 0.8127
vh=
0.9901 0.9686 0.9367 0.8964 0.8501 0.8000 0.7483 0.6966 0.6462 0.5979 0.5525
0.9964 0.9996
dIRM=
0.6711 0.7223 0.7742 0.8254 0.8738 0.9174 0.9273 0.8853 0.8378 0.7872 0.7353
0.5748 0.6217
dCPN=
0.6711 0.7223 0.7742 0.8254 0.8501 0.8000 0.7483 0.6966 0.6462 0.5979 0.5525
0.5748 0.6217
dIR=
0.8621 0.9071 0.9367 0.8964 0.8501 0.8000 0.7483 0.6966 0.6462 0.5979 0.5525
0.7613 0.8127
dTBN=
0.8621 0.9071 0.9455 0.9750 0.9858 0.9615 0.9273 0.8853 0.8378 0.7872 0.7353
0.7613 0.8127

c=
0.8501 0.9273 0.8501 0.8501 0.9273
0.8501 0.8501 0.8501 0.8501 0.8501
0.8501 0.9367 0.9367 0.8501 0.9273
0.8501 0.9858 0.9367 0.8501 0.9273

dsIRM = 0.4537
dsCPN = 0.4845
dsIR = 0.5141
dsTBN = 0.4835

b= 0.4835 0.5141 0.4845 0.4537
maxb= 0.5141
The result of user comments: IR

The output of the proposed hypothetical problem shows that IR Airline is preferable among users than the other three companies. This result is achieved by fuzzying the opinions of users which are expressed with verbal variables and are inherently ambiguous. This example shows

that fuzzy attitude is an appropriate tool for disambiguating opinions and ideas and organization can use it to obtain a more appropriate analysis of users' demands.

3. Conclusion and Recommendations

Today, the desire of organizations to be aware of the opinions, ideas and feelings of users and customers has increased significantly. Many people express their opinions, ideas and feelings in the form of verbal variables on the web. The two factors of using verbal variables and web dynamics add to the ambiguity and complexity of people's opinions and confuse managers to understand the real needs and opinions of users. Therefore, in this study, a model based on fuzzy theory was proposed to eliminate such ambiguity and increase the power of the organization to understand the ideas and feelings of users. The results of this research in the form of a hypothetical problem showed that the use of fuzzy theory removes ambiguity from the opinions and views of users and caused to make optimal decisions by members of the organization based on the real opinions of users.
According to the research findings it is recommended that in addition to paying attention to the opinions, beliefs and real feelings of users in order to improve the level of service, the method of receiving people's opinions in the form of verbal variables be limited and if verbal variables are used, decisions based on users' opinions must be made after clearing up the ambiguity with methods such as fuzzy and based on the real point of view and taking into account the hidden angles of ideas and feelings. Other researchers are also advised to implement the debugging process for other stages of data mining and web mining in future research, in addition to completing and expanding the stated materials.

References

[1] Al-Maimani, M., Salim, N., & Al-Naamany, A. M. (2014). Semantic and fuzzy aspects of opinion mining. *Journal of Theoretical & Applied Information Technology*, *63*(2).
[2] Ariyanejad, M.B., & Safakish, M.S. (2009). Multi-criteria group decision making. *Islamic Azad University- Boroujerd Branch.*

[3] Asfe, M., Jangizehi, M., Tash, M. N. S., & Yaghoubi, N. (2014). Ranking different factors influencing flight delay. *Management Science Letters, 4(7), 1397-1400.*

[4] Asqarpour, M.J. (2004). Multi-criteria decisions. *University of Tehran, IR.*

[5] Bojadziev, J., & Bojadziev, M. (2002). Fuzzy logic and its applications in management. *Ishiq, IR*

[6] Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent systems, 28(2), 15-21.*
[7] Casillas, J., & Martínez-López, F. J. (2009). Mining uncertain data with multiobjective genetic fuzzy systems to be applied in consumer behaviour modelling. *Expert Systems with Applications, 36(2), 1645-1659.*

[8] Dalal, M. K., & Zaveri, M. A. (2014). Opinion mining from online user reviews using fuzzy linguistic hedges. *Applied computational intelligence and soft computing*.
[9] Du, W., & Tan, S. (2009). An iterative reinforcement approach for fine-grained opinion mining. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (486-493).*
[10] Hong, T. P., Lee, Y. C., & Wu, M. T. (2014). An effective parallel approach for genetic-fuzzy data mining. *Expert Systems with Applications*, 41(2), 655-662.
[11] Hosseinkhani, J. (2013). Data Mining: A Quest for Concepts and Algorithms. *Yekan, IR.*

[12] Hosseinkhani, J. (2014). Web Mining: A Quest for Concepts and Algorithms. *Yekan, IR.*

[13] Jangizehi, M., Kenari, A. R., & Hosseinkhani, J (2018). Proposing a method to Choose the optimal process during an airline flight plan using a fuzzy network analysis process. *The 4th National Conference on Technology in Electrical and Computer Engineering*

[14] Jangizehi, M., Tash, M. N. S., & Yaghoubi, N. (2013). Choosing optimization process in the event of fligh plan interruption with the aid of network analysis process. *International Journal of Engineering, Science and Mathematics, 2(3), 46-56.*

[15] Jusoh, S., & Alfawareh, H. M. (2013). Applying fuzzy sets for opinion mining. *International Conference on Computer Applications Technology (ICCAT) (1-5).*
[16] Kahraman, Ch. (2015). Fuzzy multi-criteria decision making. *Simaye Danesh, IR*

[17] Lesani, H. (2011). Research methods in engineering and experimental sciences. *Qaem, IR.*

[18] Madani, Y., Erritali, M., Bengourram, J., & Sailhan, F. (2020). A multilingual fuzzy approach for classifying twitter data using fuzzy logic and semantic similarity. *Neural Computing and Applications, 32(12), 8655-8673.*
[19] Martens, D., Bruynseels, L., Baesens, B., Willekens, M., & Vanthienen, J. (2008). Predicting going concern opinion with data mining. *Decision Support Systems, 45(4), 765-777.*
[20] Nayak, M., & Dash, S. (2011). Fuzzy Decision Tree and Particle Swarm Optimization for Mining of Time Series Data. *International Journal of Computer Applications.*
[21] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis Foundations and Trends in Information Retrieval. *Decision Support Systems, 45(2), 171-179.*
[22] Rajabzade, A. (2014). Artificial intelligence and fuzzy logic. *Saffar, IR.*
[23] Seerat, B., & Azam, F. (2012). Opinion mining: Issues and challenges (a survey). *International Journal of Computer Applications, 49(9).*
[24] Su, Q., Xu, X., Guo, H., Guo, Z., Wu, X., Zhang, X., ... & Su, Z. (2008). Hidden sentiment association in chinese web opinion mining. In *Proceedings of the 17th international conference on World Wide Web (959-968).*
[25] Sudhakaran, P., Hariharan, S., Lu, J., Gamasu, R., Hamid, A., & Ilyas, M. (2013). Research directions, challenges and issues in opinion mining. *International Journal of Advanced Science and Technology, 60, 1-8.*

[26] Velvizhy, P., Pravi, A., Selvi, M., Ganapathy, S., & Kannan, A. (2020). Fuzzy-based review rating prediction in e-commerce. *International Journal of Business Intelligence and Data Mining, 17(1), 101-116.*

[28] Zadeh, L. A. (2003). From search engines to question-answering systems the need for new tools. In *The 12th IEEE International Conference on Fuzzy Systems, 2003. FUZZ'03. (Vol. 2, pp. 1107-1109).*

[29] Zadeh, L. A. (1999). From computing with numbers to computing with words. From manipulation of measurements to manipulation of perceptions. *IEEE Transactions on circuits and systems I: fundamental theory and applications, 46(1), 105-119.*

[30] Zadeh, L. A. (1997). Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy sets and systems*, *90*(2*), 111-127.*