

# Record linkage using fuzzy sets for detecting suspicious financial transactions

Sezi Çevik Onar<sup>1\*</sup>, Başar Öztayşi<sup>1</sup>, Cengiz Kahraman<sup>1</sup>

<sup>1</sup>ITU Industrial Engineering Department 34367 Macka Besiktas Istanbul, Turkey

\*cevikse@itu.edu.tr

## Abstract

Identifying suspicious financial transactions and linking relevant records is an important data related problem. An appropriate identification may improve fraud detection and international security. The main problems in this linking process are the missing data, errors in the entries or out of date entries. This study aims at developing an efficient method for identifying the suspicious transactions. This proposed methodology compares the information provided in the financial transactions with the black lists, and links similar entries. Different distance measures, including extensions of fuzzy sets can be used with the proposed method.

**Keywords:** record linkage, financial transactions, fuzzy distance, hesitant fuzzy sets, intuitionistic fuzzy sets

## 1. Introduction

In the recent years, the volume of collected data has been increased dramatically. The data are collected from different sources, such as telecommunication, health, manufacturing and finance sectors. In order to improve the benefits of using data, an effective and accurate data analysis is necessary. Integrating the data in various sources is important for an accurate and effective data analysis. On the other hand, the collected data usually has various problems such as missing data, errors in the entries or out of date entries. Especially, the entrees that are created by people have errors. The problems occur due to various reasons such as misspelling, key board errors or usage of abbreviations. One of the data related problems is to identify suspicious financial transactions, where there are usually multiple sources of data with errors and missing inputs. Identifying suspicious financial transactions is the process of identifying illegal activities, such as money laundering, evading of international sanctions or terrorist financing. The transfer of illegally earned money to the supposedly legitimate money is called money laundering. In money laundering, cash is introduced into the financial system via a method such as money transfer. In order to avoid money laundry, evading of international sanctions or terrorist financing, blacklists are prepared. The blacklists contain information related with the people and the companies involved in such activities. The banks and money transfer institutions are legally responsible for detecting such transactions.

Appropriately linking transaction data with the black list data bases improves fraud detection and international security. The data quality, size of the databases and maintaining privacy are the main challenges in this linking process. The data used in transactions usually have quality problems. The data contain errors, different variations and missing values. These variations and errors lead to vague and ambiguous matches. Therefore, exact comparison is not suitable for identifying suspicious transactions. The size of the black list databases affects the number of potential comparisons required between records. Increase in size increases the complexity of the linkage process dramatically. This creates a performance problem in the record linkage process. On the other hand, the financial transactions should usually be completed in a short time period. In the financial transactions, the privacy of the personal and financial information should be carefully protected. The objective of this study is to efficiently identify the suspicious transactions. The information provided in the financial transactions is compared with the black lists, and the linking entries are identified with the proposed methodology.

The rest of the paper is organized as follows. Section 2 describes the record linkage problem and reviews related work. Section 3 defines fuzzy sets and extensions of fuzzy. Section 4 introduces the distance measured used in this study. Section 5 presents the proposed fuzzy search mechanism for detecting suspicious transactions. Finally, last section concludes the paper by discussing the key findings and pointing at possible future research directions.

## 2. Record linkage

Schema matching, record linkage and data fusion are classified as the major tasks in data integration (Vatsalan et al., 2013). Schema matching is the process of identifying the attributes in several data base tables that contain the same information. Identifying records from different data sources that refer to the same entity is called record linkage (Durham et al., 2012). Data fusion is the process of merging pairs of records that have been identified as referring to the same entities into a single clean record.

Record linkage is important for linking data sets based on entities with or without a common identifier such as an identification number. The case task can involve differences in record shape or storage location. The record linkage problem is valid for different areas, including medical record or risk assessment, and can be

solved with different methods. Table 1 summarizes the recent fuzzy record linkage articles in the literature. The aims of these studies, the selected field and the utilized methods are given in this table.

<i>Authors</i>	<i>Aim</i>	<i>Field</i>	<i>Method</i>
Vatsalan et al. (2014)	Present a taxonomy of privacy-preserving record linkage techniques	Privacy preserving record linkage	Taxonomy
Abril et al. (2014)	Develop supervised learning using a symmetric bilinear form for record linkage	Risk assessment	Proposed new operator
Amorim et al. (2014)	Linking police and hospital records	Police and hospital records	Fuzzy matching and probabilistic record linkage
Durham et al. (2012)	Reveal the advantages and disadvantages of privacy preserving record linkage methods	Privacy preserving record linkage	Comparison of different methods
Wang and Ling (2012)	Matching patient records in multiple databases	Medical record	Fuzzy Rule Based System

Table 1: Record linkage literature

### 3. Extensions of ordinary fuzzy sets

Ordinary (type-1) fuzzy sets, introduced by Zadeh (1965), contain set of elements where each element of a set has a degree of membership (described by a membership function) and can take any value in the interval  $[0, 1]$ .

An intuitionistic fuzzy set (IFS)  $A$  can be defined as follows (Atanassov, 1999):

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x), \nu_{\tilde{A}}(x)); x \in X\}, \quad (1)$$

where  $\mu_{\tilde{A}}: X \rightarrow [0, 1]$  and  $\nu_{\tilde{A}}: X \rightarrow [0, 1]$  satisfy the condition

$$0 \leq \mu_{\tilde{A}}(x) + \nu_{\tilde{A}}(x) \leq 1 \quad (2)$$

In Eq. 3,  $\pi_{\tilde{A}}(x)$  shows the degree of uncertainty of the  $x$  element. For the ordinary fuzzy sets this degree is equal to 0.

$$\pi_{\tilde{A}}(x) = 1 - (\mu_{\tilde{A}}(x) + \nu_{\tilde{A}}(x)) \quad (3)$$

Hesitant fuzzy sets (HFS) are the extensions of fuzzy sets that reflect the hesitation in the assignment of membership degrees (Torra, 2009). There are several authors used HFS for solving various problems (Cevik

Onar et al., 2014; Yavuz et al., 2015). HFS, where  $h = h_E(x)$  denotes the set of possible membership degrees of element  $x \in X$  in  $[0, 1]$ , can be defined as follows (Torra, 2010):

$$E = \{ \langle x, h_E(x) \rangle \mid x \in X \} \quad (4)$$

$h = h_E(x)$  is a hesitant fuzzy element (HFE) (Xia and Xu, 2011).

The upper and lower bounds of  $h$  can be given as follows:

$$h^-(x) = \min h(x); \quad (5)$$

$$h^+(x) = \max h(x); \quad (6)$$

The complement of  $h$ ,  $h^c$ , is given as

$$h^c = \bigcup_{\gamma \in h} \{1 - \gamma\}; \quad (7)$$

$A_{env}(h)$  represents the envelope of  $h$ , and is an intuitionistic fuzzy set which is defined as follows:

$$A_{env}(h) = \{x, \mu(x), \nu(x)\} \quad (8)$$

where

$$\mu(x) = h^-(x) \quad (9)$$

$$\nu(x) = 1 - h^+(x) \quad (10)$$

The basic operations on HFEs where  $h$ ,  $h_1$  and  $h_2$  are HFEs can be given as follows (Zhang and Wei, 2013):

$$h^\lambda = \bigcup_{\gamma \in h} \{\gamma^\lambda\}; \quad (11)$$

$$\lambda h = \bigcup_{\gamma \in h} \{1 - (1 - \gamma)^\lambda\}; \quad (12)$$

$$h_1 \cup h_2 = \bigcup_{\gamma_1 \in h_1, \gamma_2 \in h_2} \max\{\gamma_1, \gamma_2\}; \quad (13)$$

$$h_1 \cap h_2 = \bigcup_{\gamma_1 \in h_1, \gamma_2 \in h_2} \min\{\gamma_1, \gamma_2\}; \quad (14)$$

$$h_1 \oplus h_2 = \bigcup_{\gamma_1 \in h_1, \gamma_2 \in h_2} \{\gamma_1 + \gamma_2 - \gamma_1 \gamma_2\}; \quad (15)$$

$$h_1 \otimes h_2 = \bigcup_{\gamma_1 \in h_1, \gamma_2 \in h_2} \{\gamma_1 \gamma_2\}; \quad (16)$$

where  $\lambda$  denotes a real number.

### 4. Distance Measures

In a record linkage process, first the differences among the entries should be measured. There are several distance measures. Each distance measure has some advantages. Thus, different entries in record linkage may require using different distance measures. The distance measures can be classified as crisp distance measures and fuzzy distance measures.

#### 4.1. Crisp distance measures

The general class of measuring the distance of two entries by counting the minimum number of operations required to transform one entry into the other is called edit distance. Levenshtein (1965) distance is a crisp,

edit distance measure for measuring the difference between two sequences. It shows the minimum number of single character edits required to change one entry into the other. The Levenshtein distance can be calculated as in Eq. 17.

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0 \\ \min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b} + 1_{(a_i \neq b_i)} \end{cases} & \text{otherwise} \end{cases} \quad (17)$$

For example the Levenshtein distance between "path" and "paste" is 3:

path → pash (substitution of "t" for "s")

pash → past (substitution of "h" for "t")

past → paste (insertion of "e" at the end).

The Levenshtein distance can be normalized by dividing it to the number of characters.

Jaro Winkler distance is another crisp, edit distance measure of similarity between two entries (Jaro, 1989). The higher the Jaro Winkler distance indicates higher similarity among entries. Jaro Winkler distance gives better solutions for short entries such as person names. The Jaro Winkler distance of two given entry can be calculated as follows:

$$d_j = \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3} \left( \frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right) & \text{otherwise} \end{cases} \quad (18)$$

m is the number of matching characters, t is half the number of transpositions

For example the Jaro Winkler distance among entries "Johnny" and "Jordan" can be calculated as follows:

m=3, |s<sub>1</sub>| = 6| and |s<sub>2</sub>| = 6|. The number of mismatched characters h/r, n/d, n/a and y/n, leading to t=4/2=2.

The Jaro Winkler distance = 0.33 × (3/6+3/6 + (3-2)/6) = 0.39

#### 4.2. Fuzzy distance measures

In the literature, there are many fuzzy measures proposed by various researchers. Barrenechea et al. (2013) give a summary of these fuzzy measures. The similarities of fuzzy sets are defined by fuzzy distance measures. They are commonly used for finding matching entries when users misspell words or enter in only partial words for the search. The distance can be selected based on the type of the fuzzy set.

Minkowski r-metric distance measure is one of the fuzzy distance measures. It is defined as follows:

$$d_r(x, y) = \left[ \sum_{i=1}^n |x_i - y_i|^r \right]^{1/r}, \quad r \geq 1 \quad (19)$$

where x and y are two points in an n-dimensional space with components (x<sub>i</sub>, y<sub>i</sub>) i = 1, 2 . . . . n.

Let E be a set and let A and B be two fuzzy subsets of E. The distance for discrete case can be calculated as follows:

$$d_r(A, B) = \left( \sum_{i=1}^n |\mu_A(x_i) - \mu_B(x_i)|^r \right)^{1/r} \quad r \geq 1 \quad (20)$$

Euclidean metric is the special case of Minkowski distance, where r = 2. The Euclidean distance can be calculated as follows:

$$E(A, B) = \sqrt{\sum_{x \in E} (\mu_A(x) - \mu_B(x))^2} \quad (21)$$

Hamming distance is another fuzzy distance measure where the distance is calculated by considering the absolute differences between the membership values. The Hamming distance can be calculated as in Eq. 22.

$$H(A, B) = \sum_{x \in E} |\mu_A(x) - \mu_B(x)|, \quad (22)$$

The distance for continuous case can be calculated as follows:

$$d_r(A, B) = \left( \int_{-\infty}^{+\infty} |\mu_A(x) - \mu_B(x)|^r dx \right)^{1/r} \quad r \geq 1 \quad (23)$$

There are also set theoretic based measures for measuring similarity among fuzzy sets. Jacard index is one of them. The Jacard index can be formulated as follows:

$$S_{XY} = |X \cap Y| / (|X \cap Y| + |X \cap \bar{Y}| + |\bar{X} \cap Y|) \quad (24)$$

or

$$S_{XY} = \frac{|X \cap Y|}{|X \cup Y|} \quad (25)$$

As an example distance between (1, 2, 3) and (2, 1, 3) due to the keyboard usage error can be calculated as follows. The membership values among the numbers in the keyboard are given in Table 2.

Membership values	1	2	3
1	1	0.7	0.3
2	0.7	1	0.7
3	0.3	0.7	1

Table 2 fuzzy membership values

A = {(1, 1) (2, 1) (3, 1)}

B = {(2, 0.7) (1, 0.7) (3, 1)}

Jacard index (A, B) = 2.4/3 = 0.8

Euclidean distance E (A, B) = 0.424

Hamming distance  $H(A, B) = 0.6$

#### 4.2.1. Intuitionistic fuzzy distance measures

The fuzzy distance measures are modified for the intuitionistic fuzzy sets. The Euclidean distance for the intuitionistic fuzzy sets can be formulized as follows (Atanassov, 1999):

$$E(A, B) = \sqrt{\frac{1}{2} \cdot \left( \sum_{x \in E} (\mu_A(x) - \mu_B(x))^2 + (\nu_A(x) - \nu_B(x))^2 \right)} \quad (26)$$

The Hamming distance between two intuitionistic fuzzy sets can be calculated as follows:

$$h_A(x, y) = \frac{1}{2} \cdot (|\mu_A(x) - \mu_A(y) + \nu_A(x) - \nu_A(y)|) \quad (27)$$

Table 3 represents the intuitionistic fuzzy membership values for the previous keyboard error example.

Membership values	1	2	3
1	(1,0)	(0.7, 0.3)	(0.3, 0.6)
2	(0.7,0.3)	(1,0)	(0.7,0.3)
3	(0.3, 0.6)	(0.7,0.3)	(1,0)

Table 3 intuitionistic fuzzy membership values

The intuitionistic fuzzy set based distances can be calculated as follows:

$$A = \{ \langle 1; (1, 0) \rangle, \langle 2; (1, 0) \rangle, \langle 3; (1, 0) \rangle \}$$

$$B = \{ \langle 2; (0.7, 0.3) \rangle, \langle 1; (0.7, 0.3) \rangle, \langle 3; (1, 0) \rangle \}$$

$$E(A, B) = 0.424$$

$$H(A, B) = 0.6$$

#### 4.2.2. Hesitant fuzzy distance measures

Hesitant Euclidean distance can be defined as (Xu and Xia, 2011; Rodriguez, 2014):

$$d_1(h_1, h_2) = \sqrt{\frac{1}{l} \sum_{i=1}^l |h_{1\sigma(i)} - h_{2\sigma(i)}|^2} \quad (28)$$

Zhang and Wei (2013) propose Hamming distance measure as:

$$d_1(h_1, h_2) = \frac{1}{l} \sum_{i=1}^l |h_{1\sigma(i)} - h_{2\sigma(i)}| \quad (29)$$

where  $h_1, h_2$  are HFEs and  $l$  is the number of elements in a HFE, which is called length.

The length of HFEs can be different and in some cases the values can be given without order. For such cases, first the elements should be ordered in an increasing or decreasing order, and then if thin the set should be ordered and then the different lengths of the HFEs should

be equalized by extending the one of the values until all elements have the same number of elements (Xu and Zhang, 2013). In the pessimistic approaches the minimum values are extended whereas in the optimistic approaches extend the maximum value.

The keyboard example is used again to show how distance of HFSs can be used for the record linkage problems. Table 4 represents the hesitant fuzzy membership values for afore mentioned keyboard error example.

Mem-ber-ship values	1	2	3
1	{1}	{0.7, 0.5, 0.3}	{0.3, 0.6}
2	{0.7,0.5, 0.3}	{1,0}	{0.7,0.5, 0.3}
3	{0.3, 0.6}	{0.7,0.3}	{1,0}

Table 4 Hesitant fuzzy membership values

$$A = \{(1; (1)) (2; (1)) (3; (1))\}$$

$$B = \{(2; (0.7, 0.5, 0.3)) (1; (0.7, 0.5, 0.3)) (3; (1))\}$$

$$E(A, B) = 0.277$$

$$H(A, B) = 0.33$$

### 5. Fuzzy search mechanism for detecting suspicious transactions

In this study, a fuzzy search mechanism is proposed for detecting suspicious transaction. This proposed fuzzy search mechanism enables selecting the appropriate domains with different importance degrees.

The first step of this methodology is to select the appropriate domains. In suspicious transaction detection process *name*, *surname*, *id number*, and *postal code* are commonly used domains. The second step is to assign importance weights to the domains. Linguistic importance degrees have been used to evaluate the importance of domains. The linguistic expressions are transformed to fuzzy sets for evaluation. In this process, classical fuzzy sets or their extensions can be used. Figure 1 shows an example for the fuzzy mappings of linguistic expressions. In this example, classical triangular fuzzy sets are utilized.

In the third step, the distances among the domains are calculated. In this step, depending on the characteristics of the domain, different distance measures can be used. The above mentioned measures other than Jaro Winkler, reveal the distances among records. Our objective is to calculate the similarities between the linkages. Therefore, these distances are transformed to similarities by subtracting distance measures from 1.

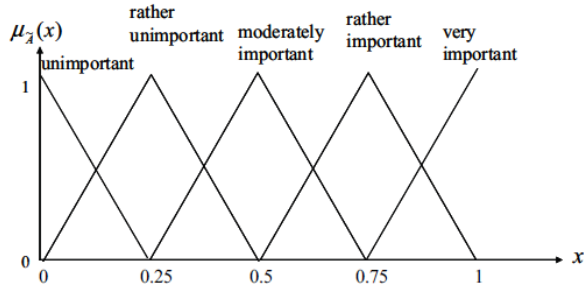


Figure 1 Membership values of the linguistic expressions

The fourth step is to calculate overall similarities of each record. In this step fuzzy weighted average is used. The following formula is applied to calculate the similarity degree of each record:

$$wa_i = f(x_1, \dots, x_n, \tilde{w}_1, \dots, \tilde{w}_n) = \frac{\sum_j^n \tilde{w}_j x_j}{\sum_j^n \tilde{w}_j} \quad (30)$$

The last step is to defuzzify the obtained scores, and listing the entries above the threshold value. In this step the threshold values are selected depending on the sensitivity of the transaction. For the high risky transactions lower threshold values are selected. The entries higher than the threshold values indicate a possible record linkage.

### 5.1. Illustrative example

The following example illustrates how the proposed record linkage mechanism can be used (see Table 5). In this example there are four domains namely, *ID*, *Name*, *Surname* and *Postal Code*. The experts assign the following importance values: *ID* domain *very important*, *name* domain *rather important*, *surname* domain *rather important* and *postal code* domain *moderately important*.

Entry	ID	Name	Surname	Postal Code
<i>initial</i>	123	Johnny	Wilson	21
1	213	Jordan	Will	12
2	123	John	Wilson	21
3	312	Johnny	Willy	12
4	313	Jordan	Wilson	21

Table 5 Illustrative example

In real life cases, the characteristics of the domain may necessitate usage of different distance methods since each method has advantages and disadvantages. In this illustrative example, the distance of each domain is calculated with different methods. The distances in *ID* domain is calculated via the fuzzy intuitionistic Euclidean distance; Jaro Winkler is used for measuring the distances in *name* domain; Levenshtein is used for measuring *surname* domain and; fuzzy hesitant Euclidean distance is used for measuring distances in *postal code* domain. The initial entry is compared with the other 4 entries, and distances and similarities between

the initial entry and these entries are calculated (see Table 6).

Entry	ID	Name	Surname	Postal Code
1	0.424	0.39	0.5	0.526
2	0	0.805	0	0
3	0.778	1	0.5	0.526
4	0.97	0.39	0	0

Table 6 Distances and Similarities of the Entries

The calculated distances are transformed into similarities by subtracting distance measures from 1 and the similarities in Table 7 are obtained.

Entry	ID	Name	Surname	Postal Code
1	0.576	0.39	0.5	0.474
2	1	0.805	1	1
3	0.222	1	0.5	0.474
4	0.03	0.39	1	1

Table 7 Similarities of the entries

The overall similarities of each record are calculated with fuzzy weighted average. Table 8 shows the overall fuzzy similarity and defuzzified score of each record.

Entry	Score	Defuzzified Score
1	(0.25,0.33,0.46)	0.340
2	(0.48,0.65,0.89)	0.661
3	(0.26,0.38,0.52)	0.384
4	(0.24,0.39,0.54)	0.392

Table 8 Weighted similarities of the entries

The threshold for this example is selected as 0.6. The second entry is higher than this threshold value. This indicates that the initial entry matches with the second entry. If the given list is a black list, the initial entry should be considered as suspicious.

## 6. Conclusion and further suggestions

In this paper, we develop a new distance based record linkage algorithm for detecting suspicious transactions. This new method enables using different measures including crisp, fuzzy, fuzzy hesitant and fuzzy intuitionistic distance measures. Different entries in record linkage require using different distance measures since each distance measure has several advantages. Using different measures for different domains is expected to improve the performance of record linkage process. The paper contributes to the literature with this new record linkage method that can be used with different data characteristics. An illustrative example has been used to show the applicability of the proposed method.

For the further studies, the performance of different measures should be evaluated and tested with a real data set. In order to reveal the robustness of the problem a sensitivity analysis of the relative importance of the domains can be performed. The framework presented in the paper is applicable in a large variety of record linkage problems. The fuzzy rule based inference systems and adaptive neuro fuzzy inference systems can be used in order to improve the record linkage performance.

## References

- [1] D. Abril, G. Navarro-Arribas, V. Torra, Improving record linkage with supervised learning for disclosure risk assessment, *Information Fusion*, 13(4): 274-284, 2012.
- [2] D. Vatsalan, P. Christen, V. S. Verykios, A taxonomy of privacy-preserving record linkage techniques, *Information Systems*, 38(6): 946-969, 2013.
- [3] E. Barrenechea, H. Bustince, J. Fernandez, D. Paternain, J.A. Sanz, Using the Choquet Integral in the Fuzzy Reasoning Method of Fuzzy Rule-Based Classification Systems, *Axioms*, 2: 208-223, 2013.
- [4] E. Durham, Y. Xue, M. Kantarcioglu, B. Malin, Quantifying the correctness, computational complexity, and security of privacy-preserving string comparators for record linkage, *Information Fusion*, 13 (4): 245-259, 2012.
- [5] K. Atanassov, *Intuitionistic Fuzzy Sets*, Springer-Verlag, Heidelberg, 1999.
- [6] L.A. Zadeh: 'Fuzzy sets', *Information and Control*, 8 (3): 338-353, 1965.
- [7] M. A., Jaro, Advances in record linkage methodology as applied to the 1985 census of Tampa Florida. *Journal of the American Statistical Association*, 84 (406): 414-20, 1989.
- [8] M. Amorim, S. Ferreira, A. Couto, A Conceptual Algorithm to Link Police and Hospital Records Based on Occurrence of Values, *Transportation Research Procedia*, 3: 224-233, 2014.
- [9] M. Yavuz, B. Oztaysi, S. Cevik Onar, C. Kahraman, Multi-criteria evaluation of alternative-fuel vehicles via a hierarchical hesitant fuzzy linguistic model, *Expert Systems with Applications*, 42(5): 2835-2848, 2015.
- [10] N. Zhang, and G. Wei. Extension of VIKOR method for decision making problem based on hesitant fuzzy set', *Applied Mathematical Modeling*, 37: 4938-4947, 2013.
- [11] R. M. Rodriguez, L. Martinez, V. Torra, Z.S. Xu, F. Herrera, Hesitant Fuzzy Sets: State of the Art and Future Directions. *International Journal of Intelligent Systems*, 29 (6): 495-524, 2014.
- [12] S. Cevik Onar, B. Oztaysi, C. Kahraman, Strategic Decision Selection Using Hesitant fuzzy TOPSIS and Interval Type-2 Fuzzy AHP: A case study, *International Journal of Computational Intelligence Systems*, 7(5): 1002-1021, 2014.
- [13] V. Torra. Hesitant fuzzy sets. *International Journal of Intelligent Systems*, 25(6):529-539, 2010.
- [14] V. Torra, Y. Narukawa, On hesitant fuzzy sets and decision, *Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on Fuzzy Systems*, 20-24:1378-1382, Aug. 2009.
- [15] V.I. Levenshtein Binary codes capable of correcting deletions, insertions, and reversals *Soviet Physics Dokl.*, 10:707-710, 1965.
- [16] X. Wang and J. Ling, Multiple valued logic approach for matching patient records in multiple databases, *Journal of Biomedical Informatics*, Volume 45(2): 224-230,2012
- [17] Z. Xu, and X. Zhang, Hesitant fuzzy multi-attribute decision making based on TOPSIS with incomplete weight information, *Knowledge-Based Systems*, 52: 53-64,2013.
- [18] Z.S. Xu, and M.M. Xia, Distance and similarity measures for hesitant fuzzy sets, *Information Sciences*, 181: 2128-2138, 2011.