Six Fuzzy Morphology Methods for Roles of Combines Standard Sentences Persian language

Haniye Rezaei[1], Homayun Motameni[2*], Behnam Barzegar[3]

Abstract–Ability to remove ambiguities is one of the main characteristics of the fuzzy systems in resolving the problems like NLP and morphology. On the other hand, most of the conducted studies in the field of morphology of Farsi language are dealing with analysis of words by means of HMM statistical method. Therefore, this paper has conducted the statistical morphology in the role of words in a sentence in the two sets of independent roles like: "Subject, adverb, possess, possessive, subject, subject header, appositive, conjunct, conjunctive, exclamation and proclaimed" using Fuzzy system. Also, in this Fuzzy system, the Max (Product) Fuzzifier by Bi-gram labeling was used. In addition, regarding the importance of defuzzifier of step one, in all fuzzy systems, for the first time, 6 types of defuzzifiers of 'Maximum membership, center of gravity, weighted average, mean of maximum, smallest of maximum, largest of maximum, center average' were implemented and obtained results have shown that the Center of gravity defuzzifier method with the mean of 63.698% had better results in comparison with other defuzzifer methods.

Keywords: Fuzzy System, Bi-gram, difuzzifire, independent roles, dependent roles.

1. Introduction

As this paper explores the labeling of vocabulary words in Farsi in the case of role of words in sentences, and because of using the Bi-gram labeling method during the proposed approach, this research can be seen in the range of labeling of words via statistical methods.

Thus, according Section 2, most studies in Persian deals with the study of words, on the other hand, statistical method proposed in this paper and previous research of phase system, used fuzzy system for decision making, that it is important to know that research in the field of statistical morphology applied HMM methods or other methods other than the fuzzy method.

Another novelty in this paper is about investigation of 6 types of difuzzifire approach in Bi-gram labeling method which has the highest output in relation to the Uni-gram labeling method and com-binational methods in past researches.

Also, it should be noted that in these calculations and fuzzy decision making, in three directions (role, type and words) Bi-gram labeling was considered. One of them is the Bi-gram labeling from analysis of the words of input word (one of the system inputs), to the words combination. The second Bi-gram labeling calculated the weight of words caused by the words forming the sentences and the final Bi-gram labeling in these calculations, is related to the words combination. After fuzzification and creation of a fuzzy relation of Max (product) among the labels, finally comparison of 6 types of defuzzification methods in Bi-gram approach was done and results were compared [1, 2].

2. Related works

The studies on lexicology as a great part of data mining are started by Petr Trojanskij,1933 [3, 4]. Later, during these years, data mining were evaluated by different methods [5]. One of the methods is Rule Base [6], statistical and probability [7], memory-based, combination, etc. [8]. Most of the studies on processing of natural languages are in two statistical sets, Rule- Base or a combination of both methods with other method as: [9] in Indian language and [10, 11] in Arabic language. The majority of combination methods are rule based and statistical methods combination. Thus, rule-based and statistics-based methods are reviewed.

Rule-based method: This method includes grammar for making sentences and making words. One of the

1 Department of Computer Engineering, Sari Branch, Islamic Azad University, Sari, Iran.Email:haniye.rezaei@gmail.com

2* Corresponding Author: Department of Computer Engineering, Sari Branch, Islamic Azad University, Sari, Iran. Email: motameni@iausari.ac.ir

3 Department of Computer Engineering, Babol Branch, Islamic Azad University, Babol, Iran.Email:barzegar.behnam@yohoo.com

Received: 2024.11.18; Accepted: 2025.02.02

researchers on using Grammar in NLP in non-Persian-Arabic languages is Chomsky, 1956. He was a pioneer in this research. In the same year, Kleene performed the grouping and prioritization of words based on rules with finite automats and regular terms [12]. As continuing the researches [13], labeling of speech was performed with finite state and [14] with simple rules and by rule-based method. Research [15] has implemented labeling part of speech by rule-based method on Indian language. Backus [16] implemented semantics by rule-based method. Also, the study [17] refers to the models with the rules based on dependence at conceptual level of web pages [18]. The researches on machine translation in Catalan-Spanish can be based on the rules of this language. Regarding Persian-Arabic, by rule-based method, Khoja and Gar-side performed root search of words by using some models of words [19]. In another method by counting pre-fixes and suffixes and rules of Arabic language [20] searched the roots of words. Taghva et al., [21] performed roots searching of Arabic words by Khoja method with the difference that no dictionary was used. The researches of Ababneh et al., [22] are regarding root search of words in Arabic to improve the results of searching. Buckwalter in the study [23] implemented morphological analysis of Arabic. Researches [24, 25] applied rule-based method to improve root search and clarification of Arabic words and [26] performed researches on the system evaluating root search of Arabic by that Grammar and the relevant rules. Bateni [27] achieved the structure of different sentences by Persian grammar and then evaluated the conversion of each sentence to another type.

Probability and statistical method: The researchers who performed some studies on language processing in non-Persian-Arabic languages as: Kaplan performed an empirical study on statistical method [28] in the target language. Also, the study [29] is the tagging of random words of 5 languages automatically and the results are tested on 7 languages. Research [30] is the tagging of English words by a probable model. Researches [31, 32] are taggers of words in speech determination as implemented based on statistical model. In addition, researches [33] have applied statistical models and methods for machine translation with the difference that [34] is on machine translation of English and Russian. In addition, to find the morphological form of words and classification of terms, by statis-tical methods researches [35, 36] are performed. In Persian-Arabic languages researches as statistical based [37], N-gram tagger is used in Arabic to find the roots of words. IN the research [38], N-gram statistical method is used to classify Arabic documents. In some researches, N-gram statistical tag-ger and Hidden Markov Model statistical method [39, 40, 41] are used to tag Arabic speech [42]. The researches on Persian by statistical method like Arabic are performed to use HMM, N-gram as statistical morphological analyzer [43] and grammar tagger of Persian vocabularies [44]. These re-searches have presented a bright path in lexicology and processing of natural language namely in Persian. It is worth to mention that in lexicology of Persian, some researchers as “Bijan Khan” و “Shams Fard” have conducted effective researches [45, 46].

After 3 decades of the first research activities in lexicology and natural language processing, fuzzy theory was presented by Lotfi A. Zadeh” in a paper called “Fuzzy Sets” [47]. Later, in 1973, fuzzy control was established. This theory was used widely in many research fields and as an expert system could have a special position in most affairs including artificial intelligence (AI), smart systems, medicine, chemical industry [48], transportation industry, etc. One of the most important applications of this system is data mining and lexicology is a part of it and has received less attention by the re-searchers. The first research on speech determination and model diagnose by fuzzy system is dedicated to [49]. To achieve web browser models in data extraction by fuzzy system, we can refer to [50]. In fuzzy decision tree in data mining, we can consider the first studies regarding [51] and then to [52, 53]. In words tagger by fuzzy networks [54] and using intuitive fuzzy logic in data mining, we can refer to research [55].

Considering all advances in fuzzy data mining it seems that (as [56] and [57]), fuzzy morphology, especially in Persian and Arabic, and in particular the combination of words in the sentence, which deals with the meaning of words in sentences, have paid less attention by the researchers [58]. However, specific applications and the broad results of morphology systems including the "machine translation [34], summarization [59], filtering [50], speech recognition, speech synthesis [60, 61], indexing and combining texts," can be pointed which makes it essential to conduct studies for research on morphology and operation phase systems.

Therefore, in this paper, using a method [1] based on fuzzy system decisions are made in the role of words in sentences in Farsi. The study has four important aspects, one because this study examines the role of words in sentences, not of different words in the sentence, which means that the results of this study refer to the concept of sentence. Second, Bi-Gram labeling method was used in a fuzzy system that puts the method in the range of statistical methods. Third, the impact of any defuzzification method in recognition of role of words in sentences was studied. The

fourth reason of im-portance of this study is that in each of examined defuzzificators, both dependent and independent roles as well as common and less common dependent roles were evaluated separately.

3. Method of determining the fuzzy role

The studied Fuzzy method considers different levels of membership for each word according to two factors of forming letters of the word, and the moment of transition from analysis to combination in the form of fuzzy values. In addition to taking the process of defuzzification, even the possibility of any role after another, is effective on the defuzzification. Therefore, discussed fuzzy computational methods using the demystify property of fuzzy systems, and taking into consideration all aspects affecting the words role [62, 63], is trying to resolve the problem and identify the role of words in sentences [64].

Therefore, Specific feature this method can be stated the following:

1. Possibility of training the expert system using the grammatical rules of grammar Persian Language, in analysis and composition.
2. Use of the word database for each role, in order to obtain the weight of each word due to the letters of the word.
3. Using statistical computing relating to the transition from analysis to combine words.
4. Using statistical computing relating to the presence of each of the other role.
5. Obtain independent and dependent separate roles.
6. Evaluating six different types of defuzzification.
7. Compare the results of the six different types of defuzzification.
8. Use fuzzification, the impact poor relations, less, and the impact strong relationships, further, be.
9. Possibility of using this method in other languages. [65, 48, 1]

The method for determining the fuzzy role has 8 steps. Therefore, in general the algorithm of fuzzy calculations has 8 steps as shown below:

1. Receiving the sentences and analysis of them from user
2. Extracting the required matrices
3. Forming the possible states according to the number of words of input sentences.
4. Removing the impossible scenarios
5. Calculating the matrix of Realation_tarkib/μ_A using the expressed terms of calculations in step 6.
6. Deriving different defuzzifires available in 4-3 section using the μ_A in 2-3-3 section and x_i of 3-1-3-3
7. Repeating steps 5 and 6 for all possible scenarios.
8. Obtaining the biggest output result of a variety of possible scenarios of phrasing and displaying the output results [66, 1].

As can be seen in Fig 1 the fuzzy system approach includes 4 overall steps. In this part, these 4 steps are described in detail to identify the role of words using fuzzy system. So initially there are descriptions for inputs required for the system, then the rules of grammar of Persian language required for training of fuzzy system for conclusion engine [67], and then steps of taking required conclusions by mean of fuzzy system are described. At the 4th step, the process of extracting results by means of different defuzzifires is descried and their steps are compared. Finally, a hypothetical example, and at the end of a hypothetical example with the four methods are described [8, 68, 69].

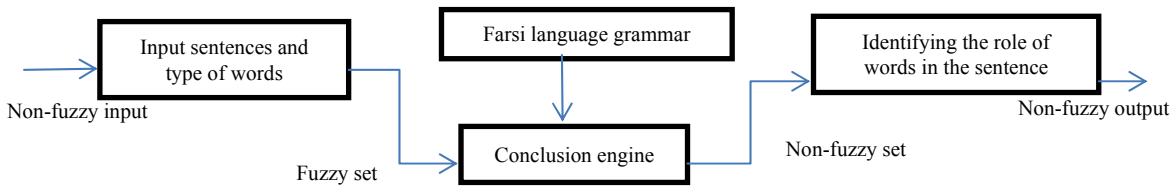

Fig 1. Fuzzy system for identifying the role of fuzzy words

3.1. Input sentences and type of words

One of the inputs of this system is the sentences in the Farsi language. It is worth noting that these sentences and words are in standard form. In this study, the slangy sentences are not considered. Since some of the Persian words can be bi-partial, (like the 'has-been' verb, it I essential to consider the roles of accurate typing rules, like treatment of half-distance standard.

Another input of system is the true tag of analysis, related to each of the words that can take advantage from most of successful researches conducted in this field to identify the type of words in the sentence. To obtain the analysis of the words in the input sentences, it is possible to use an executable file "software of natural language processing" made in "laboratory networks of Ferdowsi University of Mashhad". This software represents only Tags of analysis of input sentences clearly, and is easily accessible in the site of this laboratory. Analysis tags discussed here are included of 10 tags: (verb, noun, adverb, adjective, pronoun, preposition, including overnight ;!;). [70].

3.2Persian grammar rules

The number of possible combinations or derived from Persian grammar for this project are 194 cases, with different wording. Therefore, the statistical results of 194 different wording, is used to train the expert system. In this regard, firstly the analysis of this type of 194 wording of Persian grammar and then their combination with the help of Persian grammar books in Persian grammar school and high school with the help of experts in the field, extraction and then to train the fuzzy system are applied [1].

3.3Conclusion engine

Persian language conclusion engine by means of fuzzy system, is composed of two sections of labeling and defuzzificating. Therefore, firstly three different types of labeling required for this method are as shown in Section 3.3.1, then using such labeling, Defuzzification is described by Max (product) method [8, 71].

3.3.1. Bi-gram labeling

One of the major successful labeling methods used and managed in statistical morphology, is the labeled N-gram labeling approach. This method is classified in different classes including Uni-gram as the first floor, Bi-gram as the second floor and Tri-gram as the third floor. In general, the formula for calculating the N-gram method is the following[6, 7].

$$\sum_{i=1}^{M}\begin{pmatrix} i \dots\ i-N-2 \text{ after } i-N-1 \\ \text{after } i+N\% \text{ of repeat} \end{pmatrix} \quad (1)$$

In Eq. (1), the general state of calculation of words weight by means of N-gram method, M is the number of letters in each word or the words of a sentence, N shows the ranking of the labeling of N-gram. In a simpler word, 3 levels of N-gram labeling can be stated as shown in Table 1 [72, 71, 73].

Table 1. Comparison of 3 classes of labeling of N-gram

No.	Type of labeling	Changes of sentence 1
1	Uni-gram	$\sum_{i=1}^{M}$(i letter or word)
2	Bi-gram	$\sum_{i=1}^{M}\begin{pmatrix}\%\text{ of repeating the } i+1 \\ \text{word or letter after the word} \\ \text{or letter of i}\end{pmatrix}$
3	Tri-gram	$\sum_{i=1}^{M}\begin{pmatrix}\%\text{ of repeating the word or letter of} \\ i+2 \text{ after the word or letter of } i+1\end{pmatrix}$

In this paper, the method of Bi-gram labeling is used. The reason for this choice is that in the [1] article three different labeling were examined and the results of the comparison indicate that labeled Bi-gram method is superior. So therefore, in this article, only the Bi-gram method is required for early labeling as the fuzzy systems input.

Of course, 3 different types of Bi-gram labeling re needed, in continue the steps of the Bi-gram used in the fuzzy calculations will be discussed. Notably, all three types of labeling mentioned below were applied by Mircosoft Excel 2013 outside of this research area which is Microsoft visual studio VB.net 2012, it has been calculated and then used as tables or arrays that trains fuzzy system.

Labeling based on the letters forming the words

In this type of labeling, firstly the both database from each of the two groups, independent roles "subject, prepositions, Subject, predicate, object, complementary and verb" and related functions, " adjectives, adverbs, possess, possessive, conjunct, conjunctive, appositive, exclamation, proclaimed, the unknown and the prepositions" with three separate symbols for words such as"!;, " are available in each group.

In overall calculations, 10 roles or tags for independent roles and 15 tags for dependent roles were calculated. In this statistical work, the Bank of words Software of " analysis of Farsi sentences (Pars Process', 'The body of bi-Jen-Khan words' and for the verbs, besides them also the 'Bank of verbs of Farsi language data references, Version 3.0' was also used. For the Names also the 'bank of name culture and nomination of Iran Registration Office were used. The total number of words in the database, for all roles is 76,274 words. Then, by the help of the Microsoft

Excel software, the Bi-gram content of the forming words of 'ابپتثجچحخدذرزژسشصضطظعغفقکگلمنوهیئیةؤإأءآ' , which is composed of different letters and words, in the form of %, calculated by the help of the sentence available in the second raw of the table 1. Therefore, the phrase 1 which is in a general form, changes to the form of phrase 2 at this step.

$$\sum_{i=1}^{M}\left(\%\ of\ repeat\ of\ word\ i\ after\ the\ word\ i+1\right) \qquad (2)$$

In Eq. (2), the overall condition of calculation of words weight by means of Bi-gram method, M is the number of letters forming each input word [7, 74].

In fact, in this calculation, the total of %of presence of each letter after the other one is in each one of the roles where in each role, a 2-dimensional array of (45*45) is obtained. By summation of these values (by the help of phrase 2) in each word of input sentence, weight of words I derived based on the forming letters of each word. The output of these calculations, is a 2-dimensional array (M*10) for independent roles, where 10 is the number of roles available in this set and M is the number of words of input sentence. In this way, a two-dimensional array of (M*15) is derived for the dependent roles where 15 is the number of dependent roles and M is the number of words in input sentence. These tables or 2 dimensional matrices are called as the Matrix_B. In the first raw of the Table 2, there are some of the calculations of these labeling [72, 75, 1].

Labeling on the basis of the following words any word before

This label is another type of labeling for fuzzy processing. This labeling can consider the words of analysis, as well as the role of words in a sentence. In fact, this label checks that after any analyzing word, in place of i^{th}, in what % of cases which role of a combination locates at the location of $i+1^{th}$. In fact, it could be said that this label statistically determines the moment of decomposition transition (entry) to the combined survey. So, for the statistical calculations, 194 different wording have been expressed in section 3-2. After this calculation, the output of this stage is put in the two-dimensional matrix with dimensions (10 × 15) and (10 × 10) and classified as Bi_gram_tarkib. In these matrices there are 10 types of words in the analysis, along with three separating symbols, independent roles 10 and 15 is the number of dependent roles. So, it is noteworthy that the general term for the overall 3 is 1 at this point[6, 9, 10, 4, 7].

$$\sum_{i=1}^{M}\begin{pmatrix}\%\ \text{of repeating role i after the}\\ \text{type of i}+1\end{pmatrix} \qquad (3)$$

In Eq. (3), the general case for calculation of weight and role of words by means of Bi-gram method, M represents the number of words in the input sentence. This statement based on the two inputs of the calculations, one is kind words, and the other one is input sentence, performs the calculations. In each step, these calculations are repeated based on the assumptions available from the past and its values are derived from Bi_gram_tarkib matrix. Number of assumption conditions is changed based on the number of words in input sentence. The possible numbers for the dependent roles and independent roles are about $15^{\text{number of words}}$ and $10^{\text{number of words}}$. The second raw in the table 2 shows a sample of these types of labeling [1].

Labeling based on any role by the words after the word before

This labeling is special to the different combination conditions. The input of sentence analysis does not have any effect on the matrix calculations. In fact, this labeling is caused by statistical calculations in 194 types of phrasing of sentences extracted from Persian grammar. After obtaining the composition of these wording using Persian grammar rules, for training of fuzzy system, 2 sets of two-dimensional matrices with dimensions of (10 × 10) were 10 is the number of independent roles and (15 × 15) were 15 is the number of dependent roles are obtained by the name Bi_gram_kol. The matrix is derived by % of attendance of any role in the composition at i + 1 location after another role or repeat of the same role at the i^{th} iteration. So, the general term of 1 is changed to term 4.

$$\sum_{i=1}^{M}\begin{pmatrix}\%\ \text{of repeating role i}\\ \text{after the type of i}+1\end{pmatrix} \qquad (4)$$

In Eq. (4), the general case for calculation of weight and role of words by means of Bi-gram method, M shows the number of words in input sentence. In the third raw of Table 2, a sample of such labeling calculations is shown. Table 2, calculations for the three types of labeling in sample sentence 'باران زود بارید' /baran zud barid/ to mean 'rain fall rapidly' و 'باران' to mean 'Rain' is shown [6, 10, 4, 1].

3.3.2 Fuzzifire of μ_A=MAX (product)

The machinations use the integers in the system, and cannot use fuzzy numbers. On the other hand, in a system such as

the detection system of identifying role of words in the sentence, there are some convertors for turning the integers to the fuzzy words, so that the system is able to establish the correlation among different sections. Therefore, in this section we investigate the fuzzification method used in this study [76, 51, 8].

One of the frequently used Fuzzifire methods in fuzzy calculations is the Max (product) approach. In fact, the reason for using Fuzzifire in decision making systems is to obtain a correlation among the problem elements to include them in the calculations and removing the ambiguities, making the decision and resolving the problem. The reason of using Fuzzifire of Max (product) is that in multiplying of values it is possible to create the multiplier form analysis to combination and also the weight of words so that both of these cases have the biggest membership degree. On the other hand, because of choosing the Ma values of the multiplies the biggest value is settled in the calculations. Therefore, in this section a relation is derived between the two matrices of Matrix_B and Bi_gram_tarkib, where the output value of these calculations is fuzzy [77, 78, 79].

Table 2. 3 types of Bi-gram matrices required for the fuzzy calculation

Calculation of weights of roles	Changes of statement 2	Required input	Name of output matrix	No
% of repeating word 'b' after 'a'+ % of repeating word 'a' after 'r'+ % of repeating word 'r' after 'a'+ % of repeating word 'a' after 'n'	$\sum_{i=1}^{5}\left(\%\text{ of repeating role i} + 1\text{ after the type of i}\right)$	'باران' /baran/	Matrix_b	1
Percent repeat of role 'adverb' after the type of 'noun'+ % of repeating the role of 'verb' after the 'adjective'"	$\sum_{i=1}^{3}\left(\%\text{ of repeating role i} + 1\text{ after the type of i}\right)$	باران زود ' بارید'/baran zud barid/ and 'noun, adjective and verb'	Bi_gram_tarkib	2
Percent repeat of role 'adverb' after the type of 'Subject + % of repeating the role of 'verb' after the 'adverb	$\sum_{i=1}^{3}\left(\%\text{ of repeating role i} + 1\text{ after the type of i}\right)$	باران زود ' بارید'/baran zud barid/	Bi_gram_kol	3

$$m = \text{Number of Words}\text{،}1 < i < 10\text{،}\ 0 < j < m\text{،}1 < k < 25$$
$$Relation_{tarkib} = A \circ B = \left|t_{i,j}\right|\text{،}t_{i,j} = max_{k=1}^{m}(r_{i,k} \times s_{k,j}) \quad (5)$$

In Eq. (5), Max (product) calculations, Relation_tarkib is an output matrix derived by correlation of two matrices with two dimensions (10* Number of words), where 10 is the number of analysis tags. Two matrices of A and B required here have dimension of (25*10) and (number of words*25) where 10 is the number of analysis tags and 25 is the total number of combination tags. It should be noted that tags like prepositions and words separating characters are jointed in analysis and combination of dependent and independent and totally, there are 25 tags in a composition. In statement 5, m is the number of words in the input sentence. Therefore, correlation 5 with two matrices of Bi_gram_tarkib and Matrix_B is as shown in Eq. (6):

$$\text{Relation}_{\text{tarkib}_{i,j}} = \max_{k=1}^{25}\left(\text{Bi\_gram\_tarkib}_{i,k} \times \text{MatrixB\_tarkib}_{k,j}\right) \quad (6)$$

In Eq. (6), Max (product), calculations in Bi-gram labeling, k is the number of 25 tags in the composition and I is the number of words in the input sentence, j is the number of analysis tags.

Eq. (7), Calculations of membership degree, shows how to calculate the μ_A , where regarding the labeling characteristics of Bi-gram, the i and i+1 values of each word are added in Relation_tarkib, if the sum of these values is bigger than 1, it is considered as 1 [62, 65, 78].

$$0 < i < Number\ of\ words\ in\ the\ sentence$$
$$\mu_A(x_i) = \text{Relation}_{\text{tarkib}(i\text{،role of i word})} + \text{Relation\_tarkib(i} + 1\text{،role of i word)} \quad (7)$$

3.4 Defuzzification

After taking fuzzy calculations in section 2-3-3 and results obtained at this section, for making decision of the values in a machine, there is a need to converting them to integers [80]. So, we require some defuzzifires which can turn the fuzzy values to the integers. Therefore, because of importance of such defuzzifires and calculation of success in each method and obtaining the best possible defuzzifire,

among them 6 types of defuzzifires are investigated including: Maximum membership ،Center of gravity ، largest of max ،mean max membership ،smallest of max ، weighted average. In all of the defuzzication calculations described in this section, number of words in each sentence is =N [79, 76, 81, 82].

3.4.1. Max membership

In this method, the repeating percentage of each role after the other role (3-1-3-3) with the max membership degree (section 2-3-3) are considered as the answer. In fact x_i with the maximum value of u_A is considered as the equivalent number of the fuzzy number.

Since the membership values in this study are discrete, such calculations are applied as shown in Eq. (8).

$$z^* = \underset{\text{Related to the } (\max_{i=1}^{1-\text{Number of words}}}{(\text{Repeat \% of each role affter another})} \left(\begin{matrix} (\text{Membership} \\ \text{degree of i word} \\ \text{and i} + 1\ )) \end{matrix} \right) \quad (8)$$

In Eq. (8), Maximum membership, fuzzy values are the membership degree. In fact, values of membership degree for each word are investigated in 25 different roles ad at the biggest membership degree, x_i value is extracted for that path.

Of course, it is possible that number of maximums in this fuzzy system become more than 1. So, in such cases, the first membership degree found among the memberships is considered as the repeating percentage of that role (section 3-1-3-3) as the possibility of that composition to be happened [83, 80, 84].

3.4.2. Center of gravity

One of the defuzzifires used here is the gravity center defuzzifires, as the most common defuzzification method in fuzzy systems. Statement 9, is an overall formula for calculation of gravity center in the case of discrete degrees.

$$z^* = \frac{\sum_{i=1}^{1-\text{Number of words}} \left(\begin{matrix} \text{Membership degree} \\ \text{of the i word and } i+1 \end{matrix} \right) * \left(\begin{matrix} \text{repeat percentage of} \\ \text{each role after another} \end{matrix} \right)}{\sum_{i=1}^{1-\text{Number of words}} \begin{matrix} (\text{Membership degree of the} \\ \text{i word and i} + 1) \end{matrix}} \quad (9)$$

In the Eq. (9), Gravity center, the membership degree is the (section 2-3-3) of i^{th} and $i+1^{th}$. The calculation process of the i^{th} and $i+1^{th}$ are shown in section 2-3-3. x_i Show the repeating percentage of each role after the other (section 3-1-3-3) [83, 80, 1].

3.4.3. Largest of max (LOM)

In this method, the membership degrees (section 2-3-3) of i^{th} and $i+1^{th}$ words based on the repeating percentage of each role after another are arranged in ascending order and among them the biggest value of x_i is related to the biggest μ_A are obtained. If there is only one value of the biggest membership degree, it behaves like section 1-4-3.

Therefore, this defuzzifires can be used in the continuous and discrete systems like systems of identifying the role of words in the Farsi sentences. Statements 10 and 11, are the overall formula of calculations related to the Largest of max defuzzifires.

$$\text{hgt}(x_{i_k}) = \left\{ \left(\begin{matrix} \text{Membership degree of the i word} \\ \text{and i} + 1 \end{matrix} \right) x_i \text{of the biggest} \right\} \quad (10)$$

In Eq. (10), Supremum(μ_A), hgt is an aray where the all values of the repeating percentage of each role after the other role related to the biggest membership degrees (section 2-3-3) obtains the i^{th} and $i+1^{th}$ words or Supremum(μ_A). In fact, statement 10 is the first step of this defuzzication process. The complementary step of statement 10 is shown in statement 11.

$$Z^a = \left\{ \begin{matrix} \text{The biggest } \left(\begin{matrix} \text{repeating percentage of each} \\ \text{role after the other role} \\ \text{available in hgt} \end{matrix} \right) \end{matrix} \right\} \quad (11)$$

Eq. (11), calculation of the final output of defuzzication method of the biggest maximums, considers the biggest repeating percentage of repeating each role after another derived in section 3-1-3-3 and statement 10 in array of hgt, as the output of the defuzzication of the biggest maximum method [82, 80].

3.4.4Smallest of max (SOM)

In this method, membership degrees are arranged in an ascending order based on their x_i values (repeating percentage of each role after another, section 3-1-3-3), and among them the smallest x_i of the biggest μ_A (membership degrees of the i^{th} and $i+1^{th}$, section 2-3-3) are obtained. This method is applicable for the systems that can have a Max of μ_A, and if there is just one maximum membership degree, it behaves like section 1-4-3.

Therefore, such defuzzifires can be used in continuous systems and also discrete systems like the system of identifying the role of words in Farsi language. Statements 10 and 12 show the general formula of calculations via smallest of max defuzzication method.

Statement 10 is the first step of this defuzzication method. The complementary step of statement 10 is shown on statement 12.

$$z^b = \left\{ \begin{pmatrix} \text{The smallest} \\ \text{repeating percentage} \\ \text{of each role after another role} \\ \text{available in hgt} \end{pmatrix} \right\} \quad (12)$$

Eq. (12), calculation of the fnal output of the biggest max defuzzication method, shows the smallest values of x_i (repeating percentage of each role after another, section 3-1-3-3) obtained by statement 10 located in hgt array, as the output of the smallest min defuzzication method.
Difference of this method and the biggest max defuzzifires is in statements 11 and 12, where in statement 11, sup values are shown as the output but in statement 12, min values are considered as the output[47, 82, 83, 80].

3.4.5. Mean of max

This defuzzifire, is the mean of the x_i (repeating percentage of each role after another, section 3-1-3-3) obtained by statement 10. In fact, the mean of the x_i repeating percentage of each role after another, section 3-1-3-3)

This defuzzifire, is the mean of (repeating percentage of each role after another, section 3-1-3-3), obtained by the statement 10. In fact, it is the mean of x_i (repeating percentage of each role after another, section 3-1-3-3) with the maximum membership degree or the μ_A (membership degrees, section 2-3-3). Therefore, this defuzzicator system in the cases where the biggest μ_As is more than 1, is different from the biggest ma membership degree calculated at section 1-4-3.

This defuzzificating method uses three 11, 10 and 12 statements and finally statement 13 for obtaining the output. So, firstly by statement of 10, the x_i value with the biggest membership degree are located in htg array and then by statement of 12, the biggest x_i available in htg mentioned in statement 11, locates in the z^b. finally, to obtain the mean of maximums, statement 13 is used.

$$z^* = \frac{(z^a + z^b)}{2} \quad (13)$$

In Eq. (13), Calculation for the final output of defuzzication method of the mean of maximums, z^a is the biggest x_i (repeating percentage of each role after another, section 3-1-3-3), related to the maximum of μ_A s (membership degrees of ith and i+1th words, section 2-3-3) and z^b is the smallest maximum of μ_A s andz^*shows the output of the maximum mean defuzzication method [83, 80, 62, 5].

3.4.6Weighted average

This defuzzification method is obtained by multiplying the membership degree in mean of x_i s (repeating percentage of each role after another, section 3-1-3-3) related to the membership degree of ith and i+1th words (μ_A, section 2-3-3), divided by the sum of the mean of μ_A. Therefore, this mean has weight, because it's obtained based on the weight of their $\bar{x}_i$. Statement 14, is the overall formula of the mean weighted defuzzification method.

$$a = \frac{\sum_{i=1}^{1-\text{Number of words}} (\text{Membership degree of the i word and i} + 1) * (\overline{x_i})}{\sum_{i=1}^{1-\text{Number of words}} (\overline{x_i})} \quad (14)$$

In Eq. (14), Final calculation of the weighted mean defuzzication method, a is the final output of this defuzzifire, and$\bar{x}_i$s are the weighted means of the repeating percentage of each role after another defined in section 3-1-3-3, and the membership degree of the ith and i+1th words as shown in section 2-3-3 [47, 82, 83, 80].

3.5. An example of system of identifying the role of words in Persian sentences using fuzzy method.

In this section, using the mentioned 8-steps algorithm in section 3 and descriptions of 1-3, 2-3, 3-3 and 4-3, for example required calculations for the input sentence are shown by 'باران زود بارید./baran zud barid/ and analysis of the input 'باران'/Baran/: name, 'زود'/zud: adjective, and 'بارید' /barid/: verb'.

Firstly, the repeating percentages of each letter in different roles are obtained regarding the section 1-3. Therefore, in such calculations some tables like table 3 are used for calculation of words weight in the role of each word. In table 3, rows shows the ith letter and columns show the i+1th word.

Table 3. Some parts of precense percentage of each word after another for the role of the verb

	ا	ب	پ	ت	ث	ج
ا	0.02	4.23	0.21	1.26	0.25	0.89
ب	11.81	0.47	0.00	0.91	0.00	0.04
پ	13.46	0.00	0.00	0.00	0.00	0.00
ت	5.78	1.27	0.10	0.20	0.10	1.08
ث	12.50	5.56	0.00	1.39	0.00	0.00
ج	14.37	0.74	0.00	1.19	0.00	0.00

In table 4, the lonely part of Matrix_B array or the sum of the values for the hypothetical sentence of ‘ باران زود باريد.’/baran zud barid/ are shown in the hypothetical roles of ‘Noun, Adjective, Verb’. These calculations are totally described in section 1-1-3-3.

Table 4. Matrix_B calculations for three roles of ‘Subject, adverb and verb for example

Input word	Role subject	% of presence of each word after another in each role title	Sum of weight of role
‘باران’/Baran/	Subject	"b "Then" a+" "a "Then" r+" "a "Then" n"=	0.132+ 0.098+ 0.0115+ 0.174=1.407
‘زود’/Zud/	Adverb	"z "Then" u+" +"u "Then" d"=	0.066 0.272=0.338
‘باريد’/barid/	Verb	"b "Then" a+" "a "Then" r+" "r "Then" i+" ="i "Then" d+"	0.118+ 0.13+ 0.097 0.095=0.44

Therefore, for the hypothetical roles of ‘Subject, adverb and verb’ values of Matrix_B of this sentence are ‘باران’/baran/ →subject=1.407, ‘زود’/zud/→ adverb=0.338,’باريد’ /barid/ →verb=0.44.

For example, table 5 shows some parts of Bi_gram_tarkib matrix for the independent roles. The point that after each word in the i^{th} location which role is placed at the $i+1^{th}$ location is defined by Bi_gram_tarkib matrix.

Table 5. Some parts of Bi-gram-tarkib of independent roles

Type / role	Subject (A)	Subject headers (C)	Predicate (D)	Object (i)	Complem_ntarity (j)
Verb	3.704	0.000	0.000	7.407	5.556
noun	8.261	0.435	5.652	12.174	8.261
adverb	2.564	2.564	10.256	10.256	0.000
adjective	18.750	6.250	4.167	2.083	2.083
pronoun	4.225	0.000	7.042	14.085	4.225
letter	21.387	0.578	2.312	6.936	36.416
clause	45.455	0.000	0.000	0.000	9.091
؛	29.412	0.000	29.412	23.529	5.882
!	0.000	0.000	0.000	0.000	0.000
،	0.000	0.000	0.000	0.000	0.000

In Table 6 Results of calculations of statement 6 are shown in the input sentence of ‘باران زود باريد.’/baran zud barid/. Regarding the tables 5 and 6 and statements of 5 and 6 available in section 2-3-3, table 6 can be obtained.

Table 6. Values of Relation_tarkib matrix in the sentence of ‘**باران زود باريد**.’/baran zud barid/.

Type of words	verb	noun	adverb	adjective	pronoun	letter	Claus	!	؛	،
‘باران’/Baran/	0.0269	0.29	0.41	0.458	0.442	0.56	0.545	0.117	1	0.117
‘زود’/Zud/	0.0269	0.29	0.41	0.458	0.442	0.56	0.545	0.117	1	0.117
‘باريد’/Barid/	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

In table 6, joint roles like verb are jointed in analysis and combination and do not require calculations, therefore they have a value of Zero.

For example, table 7 shows some parts of Bi_gram_kol matrix among two matrices of Bi_gram_kol shown in section 3-1-3-3. Total Bi_gram_kol shows the presence percentage of each role after another at the i^{th} and $i+1^{th}$ location stated by statistical calculations defined at section 3-1-3-3. Bi_gram_kol is considered as the x_i in fuzzy calculations.

Table 8 shows types of calculations for defuzzifires for the hypothetical sentences of ‘باران زود باريد.’/baran zud barid/ by the type of words of ‘noun, adjective, verb’ and just one hypothetical condition of words ‘subject, adverb, verb’ regarding the statement of 14-8 and tables of 7 and 8.

Table 7. Some parts of Bi_gram_kol matrix

Role	verb	adverb	Complementarity	object	predicate	Subject headers	subject
subject	0.201	0.04	0.04	0.114	0.0	0.0	0.148
Subject headers	0	0	0.0	0.107	0.607	0.036	0.0
predicate	0.6	0.033	0.0	0.0	0.133	0.0	0.0
object	0.235	0.014	0.029	0.279	0.0	0.0	0.0
Complementarity	0.576	0.01	0.196	0.043	0.22	0.0	0.0
adverb	0.384	0.127	0	0.115	0	0.038	0.077
verb	0	0	0	0.108	0	0	0.027

Table 8. Calculations of gravity center for three hypothetical conditions

Title of defuzzifires	Calculation Sample
Max membership	$hgt(x_i) = max\big((0.29 + 0.458).(0.458 + 0.000)\big) = \{0.748\}$ $Z^* = \{0.04\}$
Center of Gravity	$Z^* = \frac{(((0.29 + 0.458) \times 0.04) + ((0.458 + 0.000) \times 0.384))}{0.04 + 0.384} = 0.485$
Largest of max	$hgt(x_{i_k}) = \max of \big((0.29 + 0.458).(0.458 + 0.000)\big) = \{0.748\}$ $Z^* = 0.04$
Smallest of max	$hgt(x_{i_k}) = \max of \big((0.29 + 0.458).(0.458 + 0.000)\big) = \{0.748\}$ $Z^* = 0.04$
Mean of max	$hgt(x_{i_k}) = \max of \big((0.29 + 0.458).(0.458 + 0.000)\big) = \{0.748\}$ $Z^a = \{0.04\}$, $Z^b = \{0.04\}$ $Z^* = \frac{0.04 + 0.04}{2} = 0.04$
Weighted average	$Z^* = \frac{(((0.29 + 0.458) \times 0.04) + ((0.458 + 0.000) \times 0.384))}{0.04 + 0.384} = 0.485$

In all methods mentioned in Table 8, finally among $10^{\text{Number of words}}$ and $15^{\text{Number of words}}$ in dependent and independent roles, the condition of arrangement with the maximum value of defuzzication is considered as the output[79, 83, 80, 5, 1].

4. Comparison and Evaluation

To test each method of 6 defuzzification methods described in Section 3.4, 70 standard random sentences are extracted from text books and websites, and used. In addition, analysis labels of 70 input sentences are used as the next input. Therefore, in this section obtained results of success percentage of each defuzzification method are compared and evaluated.

In general, there are two kinds of roles in Persian language sentences, also it should be noted that there are some overlapping roles in Persian or jointed roles among the analyzing and combining, such as "verb", or some roles expressed in various sources under different names. For example, subject or Subject headers are considered by this phrase n some resources, but also are considered as the subject in other resources. The results of analyzed roles in this study, regardless of role separation, included 5 independent and 11 dependent roles. In this study, the expression of other roles are avoided, because of the overlapping between the analysis and combine or separate characters.

Main roles including Subject (and subject headers), predicate, object and complementarity

Dependent roles including: The noun which is depended on the adjective (jointed analysis and composition) possess and possessive, adverb, appositive, conjunct, conjunctive, proclaimed and exclamation [85, 1].

$$\text{Success Rate} = \frac{(Currect * 100)}{count\ of\ Rule\ in\ 70\ sentences} \tag{15}$$

Eq. (15) Calculations of the success value in each role, calculates the amount of success percentage (Success Rate) where correct is the number of true integers extracted by fuzzy system and Count of Rule is the total number of attendance in this role during 70 sentences.

$$\text{Average success rate for the dependent}/independent = \frac{(\sum_1^{11\ \text{OR}\ 5} \text{Success Rate})}{5 \text{ or } 11} \tag{16}$$

For calculation of overall success values of two sets of dependent and independent roles, Eq. (16) is applied and Success rate I also determined by Eq. (15) for each role and then, for each set, success value can be obtained by means of Eq. (16). In this statement, 5 is the number of independent roles and 11 is the number of dependent roles. Of course, 5 and 11are calculated without considering the jointed roles, prepositions and separating characters.

$$average\ Success\ Rate = \frac{\begin{pmatrix}\text{Average success rate for the dependent} \\ + \\ \text{Average success rate for the independent}\end{pmatrix}}{2} \tag{17}$$

In each of the mentioned defuzzification methods in section 4-3, the mean of the overall success in each set can be derived by Eq. (17), where Average Success Rate for the Dependent is the success rate of roles calculated for all 5 independent roles and Average Success Rate for the Independent is the success rate of all 11 dependent roles.

Table 9. Mean of success rate percentage of each one of the defuzzifications

Title of Defuzzifications	Mean of defuzzifications
1.Max of Membership	43.648
2.Center of Gravity	63.698
3.Largest of Max	58.007
4.Smallest of Max	56.711
5.Mean of Max	53.475
6.Weighted Average	38.607

In table 9, the exact value of the mean of success rate in each of the defuzzification methods of section 4-3 is shown, both for the dependent and independent roles. Obviously, the maximum success rate in acquisition of roles of 70 input experimental sentences is for the defuzzifires of Center of Gravity, and after that for Largest of Max, Smallest of Max ،Mean of max ،Max of Membership, respectively. Finally, the minimum success was obtained by the Weighted Average method.

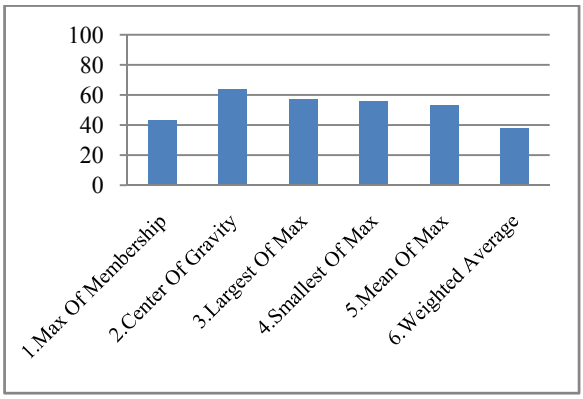

Fig 2. Comparison of the total success percentage of each one of the defuzzicators

Figure 2 shows the image of success of each one of the defuzzification methods in identification of words combination in Persian sentences. Obviously, there is a great difference among the Largest of Max and Weighted Average. But there is no difference among the Center of Gravity, Largest of Max and Smallest Max, Mean Of max, while the maximum difference among these 4 defuzzification methods is about 10%.

Table 10. Mean of success percentage of each one of the defuzzifires separated by sets of dependent and independent roles

Title of defuzzifier	Mean of independent roles	Mean of dependent roles
1.Max of Membership	56.026	31.270
2.Center of Gravity	60.843	66.553
3.Largest of Max	55.576	60.438
4.Smallest of Max	54.148	59.275
5.Mean of Max	52.288	54.663
6.Weighted Average	32.284	44.930

Table 10 shows the % of success for the dependent and independent roles, separated for each of the defuzzifires. Obviously, in the independent roles, three better positions belonged to Center of Gravity ،Max of Membership defuzzifire methods and with the little difference, is the Largest of Max. While for the dependent roles, the Max of Membership has the least percentage of success and three better positions are belonged to Center of Gravity ،Largest of Max and Smallest of Max.

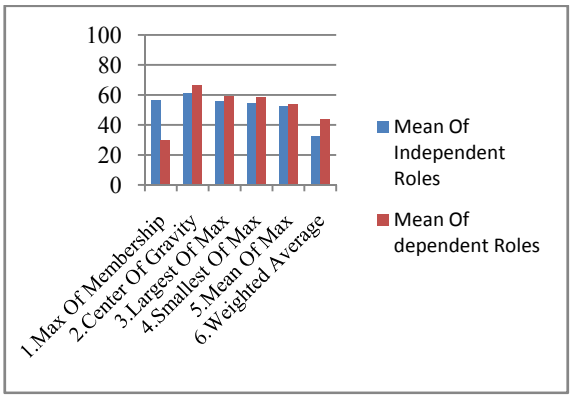

Fig 3. Comparison of the mean of success percentage for each one of the defuzzicators separated by the sets of dependent and independent roles

As shown in figure 3, difference of success percentage of dependent and independent roles in 4 types of defuzzifires are closed in two sets of dependent and independent roles. These roles with close success percentage of dependent and independent roles are included of: Center of Gravity ، Largest of Max ،Smallest of Max and Mean of Max. Just in two cases of Weighted average and Max of Membership the values of success percentage mean have more fluctuations. Also, in these two cases with high fluctuations in mean of success percentage of dependent and independent roles, the least mean of success percentage is include of the composition roles, too.

Table 11. Mean of success percentage of each one of the defuzzificators separated by the independent roles

Defuzzifires / Role	max of membership	center of gravity	mean max	largest of max	smallest of max	weighted average
Subject	55.814	6.977	41.860	46.512	51.163	81.395
Subject header	59.459	75.676	56.757	56.757	54.054	18.919
predicate	51.220	82.927	65.854	60.976	60.976	7.317

object	63.636	63.636	63.636	63.636	54.545	45.455
Complementary	50.000	75.000	33.333	50.000	50.000	8.333

Table 11 shows the amount of success for each one of the defuzzifires in each of the independent roles of Farsi language, separately. Obviously, the minimum and maximum values of success are observed in weighted average and center of gravity methods. On the other hands, there are severe fluctuations in success of identification of independent roles, center of gravity (the best method for identification of these roles) and weighted average (the worse method for identification of roles).

While the values of success with low fluctuations caused that these 4 methods become successful in identification of each one of the independent roles and also the relative success is between "50-60" as shown in table 11.

As shown in Figure 4, severe fluctuations of success in identification of each of the roles related to the two approaches of center of gravity and weighted average are really sensible. On the other hands, the mild slope of changes I the 4 remained methods are also tangible.

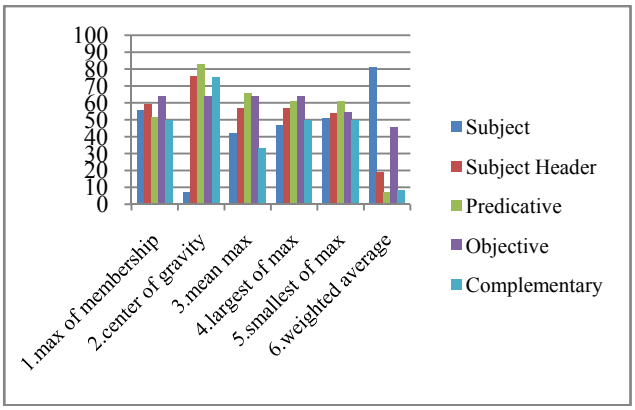

Fig 4. Comparison of percentage mean of success in each one of the defuzzicators separated by independent roles.

Obviously, center of gravity method has the biggest mean and is just usable for identification of role of the subject, with less efficiency than the other methods. Also, in Farsi language grammar, both roles of subject and subject header can be considered as the 'subject'. Therefore, in this way the center of gravity method acts so better than other methods for identification of independent role of sentences.

Table 12. Mean of the success percentage in each of the defuzzification methods separated by dependent roles

Defuzzifires / Role	max of membership	center of gravity	mean max	largest of max	smallest of max	weighted average
Adjective	10.000	55.000	10.000	70.000	10.000	35.000
Adverb	7.407	92.593	92.593	92.593	92.593	70.370
Unknown	3.846	23.077	50.000	55.769	51.923	76.923
Subject	16.667	50.000	8.333	25.000	16.667	33.333
possessive	11.765	41.176	29.412	17.647	29.412	17.647
possess	14.286	28.571	14.286	7.143	21.429	14.286
appositive	66.667	66.667	33.333	33.333	66.667	33.333
conjunct	50.000	100.000	100.000	100.000	100.000	100.000
conjunctive	50.000	100.000	100.000	100.000	100.000	75.000
exclamation	50.000	75.000	75.000	75.000	75.000	0.000
proclaimed	50.000	100.000	75.000	75.000	75.000	25.000

Table 12 shows the success percentage of each of the dependent roles in each one of the 6 defuzzifire methods in detail. It is to be noted that 'unknown' means the words without any dependent roles in the sentence. Dependent roles in reality are the roles where their existence in the sentence is depended on another role.

While investigation of the most successful dependent methods, the center of gravity defuzzifire method had success lower than 40% just in two cases of roles, while in other methods, values less than 40% were seen at least in 4 cases.

Of course, it should be noted that these two cases of minimum role in the center of gravity defuzzification method lay in the 6 first rows of the Table 13, which means a notable effect in the sentences of Farsi language.

In the case of investigation of the less successful methods for identification of dependent roles, it should be noted that the Max of Membership defuzzification method in 5 roles had a low effectiveness in the sentence of Farsi language, which means 'appositive, conjunct, conjunctive,

For investigation of the most unsuccessful methods for identification of dependent role, it should be noted that the

Max of Membership defuzzifire method had the success values of 50-66% in the 5 low-presence roles of appositive, conjunct, conjunctive, exclamation and proclaimed. It is happening while these roles have a little presence in the Farsi language sentences. Like the role of 'alternative' that because of its rare attendance in sentences of Farsi language and coverage of this role by the others, it is not considered in this study. But more importantly, in all significant 6 roles like 'adjective, dependent adverb, unknown, subject, possessive and possess' success percentage was lower than 15%.

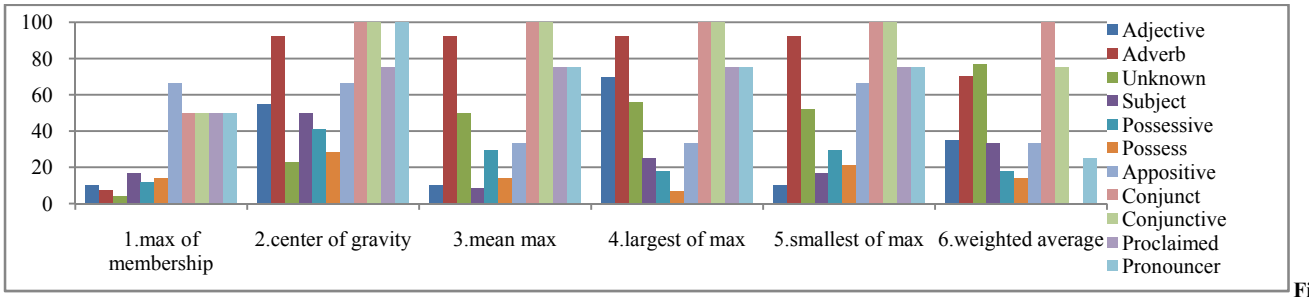

Fig 5. Comparison of success percentages for each one of the defuzzifires separated by dependent roles.

As shown in Figure 5, mean of the roles of dependent roles of sentences in Farsi language in 6 first roles, which are more important in relation to the net 5 roles, the success belongs to center of gravity ،largest of max ،weighted average ،smallest of max ،mean max and finally the Max Of Membership, while in the case of latter 5 roles of table 12, the insignificant roles, figure 5 shows that successes of center of gravity ،smallest of max ،largest of max and mean max are equal and after them there is Max Of Membership and finally there is weighted average [1].

5. Conclusion and Suggestions

By investigation of fuzzy identification of role of words, in sentences of Farsi language, it can be found that from the defects of this method and other statistical based methods is the heavy load of calculations. On the other hand, the investigation method of this work has advantages like non-dependence to the vocabulary bank in fuzzy calculations and suitable accuracy in determination of role of words in Farsi language.

Regarding the mentioned issues, in an overall mean, the Center of Gravity is the best method of defuzzification for identification of role of words in the Farsi sentences, using the Bi-Gram labeling method among other 5 methods of defuzzification.

Of course, in distinct investigation of both sets of dependent and independent roles, the Center of Gravity method is superior. In this way the second and third ranks of independent roles belong to the Max of Membership and Largest of Max method. But in the frequently-used dependent roles, second and third ranks are belonged to the Largest of Max and Weighted of Average, and for the low-used dependent roles it is the Smallest of Max and jointly at the third rank there are the Mean of Max and Largest of Max methods.

Therefore, to conduct better researches, in the fuzzy identification of role of words in the sentence and also boosting the discussed method, mentioned points can be considered:

- Completing the statistical calculations and obtained matrices of grammar, for training of fuzzy system
- Studying the fuzzy identification system and role of words in sentences of other languages.
- Studying labeling impact and also N-Gram with degrees of Higher N in fuzzy identification of role of words in the sentences
- Boosting the fuzzy system educational grammar regarding the Farsi Grammar
- Combining the fuzzy based statistical method, identifying the role of words with other methods for identifying the type of words
- Studying the non-standard and slangy sentences
- Studying the role of overlapping and similar form words, specifically.

References

[1] H. Motameni and A. Peykar, "Morphology of Compounds as Standard Words in Persian through Hidden Markov Model and Fuzzy Method," Journal of Intelligent & Fuzzy Systems, vol. 30, no. 3, pp. 1567-

1580, 2016.

[2] T. Chadza, K. G. Kyriakopoulos, S. Lambotharan, Analysis of hidden markov model learning algorithms for the detection and prediction of multi-stage network attacks, Future generation computer systems 108 (2020) 636–649.

[3] J. Wettig, S. Hiltunen and R. Yangarber, "Hidden Markov Models for induction of morphological structure of natural language," Department of Computer Science,University of Helsinki, Finland, Helsinki, 2010.

[4] S. Naderi Parizi, "Implementation of hidden Markov models associated with the ability to apply the language, grammar, search methods and the applicability of the model," Amirkabir University of Technology, Department of Computer Engineering and Information Technology, Tehran, 2007.

[5] C. C. Aggarwal, Data Mining, Switzerland: Springer International, 2015.

[6] M. Mohseni and B. Minaei-bidgoli, "A Persian Part-Of-Speech Tagger Based on Morphological Analysis," in Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta, 2010.

[7] W. Khan, A. Daud, K. Khan, J. A. Nasir, M. Basheri, N. Aljohani, F. S. Alotaibi, Part of speech tagging in urdu: Comparison of machine and deep learning approaches, IEEE Access 7 (2019) 38918–38936.

[8] P. Koehn, Statistical Machine Translation, New York: United States of America by Cambridge University Press, 2010, pp. 181-212.

[9] M. Shrivastava, "Hindi POS Tagger Using Naive Stemming : Harnessing Morphological Information Without Extensive Linguistic Knowledge," in ICON-2008:6th International Conference on Natural Language Processing,Macmillan Publishers., India, 2008.

[10] M. Bahrani, H. Sameti, N. Hafezi and S. Momtazi, "A New Word Clustering Method for Building N-Gram Language Models in Continuous Speech Recognition Systems," in New Frontiers in Applied Artificial Intelligence, 21st International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE, Wroclaw, Poland, 2008.

[11] A. F. Alajmi, E. M. Saad and M. H. Awadalla, "Hidden markov model based Arabic morphological analyzer," International Journal of Computer Engineering Research, vol. 2, no. 2, pp. 28-33, 2011.

[12] D. Dutta, S. Halder, T. Gayen, Intelligent part of speech tagger for hindi, Procedia Computer Science 218 (2023) 604–611

[13] T.-L. Tseng, F. Jiang, Y. Kwon, Hybrid type ii fuzzy system & data mining approach for surface finish, Journal of Computational Design and Engineering 2 (3) (2015) 137–147.

[14] A. Chiche, B. Yitagesu, Part of speech tagging: a systematic review of deep learning and machine learning approaches, Journal of Big Data 9 (1) (2022) 1–25.

[15] D. Modi and N. Nain, "Part-of-Speech Tagging of Hindi Corpus Using Rule-Based Method," Proceedings of the International Conference on Recent Cognizance in Wireless Communication & Image Processing, Vols. 10.1007/978-81-322-2638-3, no. 28, pp. 241-247, 2016.

[16] J. S. Rohl, "A note on Backus Naur form," Department of Computer Science, The University, Manchester 13, 2010.

[17] S. Poria, E. Cambria, G. Winterstein and G.-B. Huang, "Sentic patterns: Dependency based rules for concept-level sentiment analysis.," Knowledge-Based Systems., 2014.

[18] M. R. Costa-Juss`a, M. Farr´us, J. B. Mari˜no and J. A. Fonollosa, "Study and comparison of rule-based and statistical catalan-spanish machine translation systems," Computing and Informatics, vol. 31, pp. 245-270, 2012.

[19] A. Alnaied, M. Elbendak, A. Bulbul, An intelligent use of stemmer and morphology analysis for arabic information retrieval, Egyptian Informatics Journal 21 (4) (2020) 209–217.

[20] K. Darwish, "Building a shallow morphological analyzer in one day.," in ACL-02 Workshop on Computational Approaches to Semitic Languages, Philadelphia, PA, 2002.

[21] K. Taghva, R. Elkhoury and J. S Coombs, "Arabic stemming without a root dictionary.," in Information Technology: Coding and Computing.ITCC 2005, 2005.

[22] A. Mohamad, A.-S. Riyad and K. Ghass, "Building an Effective Rule-Based Light Stemmer for Arabic Language to Improve Search Effectiveness.," International Arab Journal of Information Technology (IAJIT), vol. 9, no. 4, pp. 368-372, 2012.

[23] T. Buckwalter, "Buckwalter Arabic Morphological Analyzer.," the Linguistic Data Consortium,, Pennsylvania, 2002.

[24] H. K. Al Ameed, S. O. Al Ketbi, A. A. Al Kaabi, K. S. Al Shebli, N. F. Al Shamsi, N. H. Al Nuaimi and S. S. Al Muhairi, "Arabic light stemmer: anew enhanced

approach," in The Second International Conference on Innovations in Information Technology (IIT'05), 2005.

[25] L. Larkey, L. Ballesteros and M. Connel, "Improving Stemming for Arabic Information Retrieval: Light Stemming and Co-occurrence Analysis.," in 25th annual international ACM SIGIR conference on Research and development in information retrieval, 2002.

[26] A. El-Hajar, M. Hajar and K. Zreik, "A System for Evaluation of Arabic Root Extraction Methods.," in fifth international Conference on Internet and Web Applications and Services., 2010.

[27] H.Alshalabi,S.Tiun,N.Omar,F.N.AL-Aswadi,K.A.Alezabi,Arabiclight-basedstemmer usingnewrules,JournalofKingSaudUniversity-ComputerandInformationSciences34(9)(2022) 6635–6642

[28] M. F. Kabir, K. Abdullah-Al-Mamun, M. N. Huda, Deep learning based parts of speech tagger for bengali, in: 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), IEEE, 2016, pp. 26–29.

[29] K. K. Zin, N. Thein, Hidden markov model with rule based approach for part of speech tagging of myanmar language, in: Proceedings of 3rd International Conference on Communications and Information, 2009, pp. 123–128.

[30] F. Pisceldo, M. Adriani, R. Manurung, Probabilistic part of speech tagging for bahasa indonesia, in: Third international MALINDO workshop, 2009, pp. 1–6.

[31] M. Attia, Y. Samih, A. Elkahky, H. Mubarak, A. Abdelali, K. Darwish, Pos tagging for improving code-switching identification in arabic, in: Proceedings of the Fourth Arabic Natural Language Processing Workshop, 2019, pp. 18–29.

[32] F. Pisceldo, M. Adriani, R. Manurung, Probabilistic part of speech tagging for bahasa indonesia, in: Third international MALINDO workshop, 2009, pp. 1–6.

[33] M. Febryanto, I. Sulyaningsih, A. A. Zhafirah, Analysis of translation techniques and quality of translated terms of mechanical engineering in accredited national journals, Professional Journal of English Education 1 (2021) 116–119.

[34] A. Xv, Russian-english bidirectional machine translation system, in: Proceedings of the Fifth Con- ference on Machine Translation, 2020, pp. 320–325.

[35] H. Aldarmaki, A. Ullah, S. Ram, N. Zaki, Unsupervised automatic speech recognition: A review, Speech Communication 139 (2022) 76–91.

[36] M. Creutz, "Unsupervised segmentation of words using prior distributions of morph length and frequency," in Proc. 41st Meeting of ACL, Sapporo,Japan, 2003.

[37] F. Ahmed and A. Nürnberger, "N-grams Conflation Approach for Arabic," in ACM SIGIR Conference, Amsterdam, 2007.

[38] A. Y. Muaad, G. H. Kumar, J. Hanumanthappa, J. B. Benifa, M. N. Mourya, C. Chola,

M. Pramodha, R. Bhairava, An effective approach for arabic document classification using machine learning, Global Transitions Proceedings 3 (1) (2022) 267–271.

[39] K. Tnaji, K. Bouzoubaa, S. L. Aouragh, A light arabic pos tagger using a hybrid approach, in: Digital Technologies and Applications: Proceedings of ICDTA 21, Fez, Morocco, Springer, 2021, pp. 199–208.

[40] M. El-Hadj, A.-S. IA and A.-A. AM, "Arabic Part of Speech Tagging Using the Sentence Structure.," in 2nd international Conference on Arabic Language Resources & Tools, Cairo, 2009.

[41] H. Hassani, Part of speech tagging (post) of a low-resource language using another language (devel- oping a pos-tagged lexicon for kurdish (sorani) using a tagged persian (farsi) corpus), CoRR (2022) abs/2201.12793.

[42] S. Alqrainy, M. Alawairdhi, Towards developing a comprehensive tag set for the arabic language, Journal of Intelligent Systems 30 (1) (2020) 287–296.

[43] H. Motameni, A. Ebrahimnejad, J. Vahidi, et al., Morphology of composition functions in persian sentences through a newly proposed classified fuzzy method and center of gravity defuzzification method, Journal of Intelligent & Fuzzy Systems 36 (6) (2019) 5463–5473.

[44] S. M. Assi and M. Haji Abdolhosseini, "Grammatical tagging of a Farsi Corpus.," International Journal of Corpus Linguistics., vol. 5, no. 1, pp. 69-81, 2000.

[45] M. Bijankhan, J. Sheykhzadegan, M. Bahrani and M. Ghayoomi, "Lessons from Building a

Persian Written Corpus: Peykare," Language Resources and Evaluation, vol. 45, pp. 143-164, 2011.

[46] M. Shamsfard, H. Sadat Jafari and M. Ilbe, "STeP-1: A Set of Fundamental Tools for Persian Text Processing.," in LREC 2010, Valletta, Malt, 2010.

[47] H. T.-P., L. K.-Y. and W. S.-L., "Mining linguistic browsing patterns in the world wide web," Soft Computing, vol. 5, pp. 329-336, 2002.

[48] F. M. Zanzotto, L. Dell'Arciprete, A. Moschitti, Efficient graph kernels for textual entailment recognition, Fundamenta Informaticae 107 (2-3) (2011) 199–222.

[49] N. Passalis, J. Raitoharju, A. Tefas, M. Gabbouj, Efficient adaptive inference for deep convolutional neural networks using hierarchical early exits, Pattern Recognition 105 (2020) 107346.

[50] Z. Elaggoune, R. Maamri, I. Boussebough, A fuzzy agent approach for smart data extraction in big data environments, Journal of King Saud University-Computer and Information Sciences 32 (4) (2020) 465–478.

[51] M. E. Cintra, M. C. Monard, H. A. Camargo, A fuzzy decision tree algorithm based on c4. 5, Mathware & Soft Computing 20 (1) (2013) 56–62.

[52] X. Sun, L. Yuan, M. Liu, S. Liang, D. Li, L. Liu, Quantitative estimation for the impact of mining activities on vegetation phenology and identifying its controlling factors from sentinel-2 time series, International Journal of Applied Earth Observation and Geoinformation 111 (2022) 102814.

[53] X. Bai, Y. Yang, Fuzzy decision tree algorithm based on feature value's class contribution level, Iranian Journal of Fuzzy Systems 19 (4) (2022) 73–88.

[54] S. Sayami, S. Shakya, Nepali pos tagging using deep learning approaches, NU. International Journal of Science 17 (2) (2020) 69–84.

[55] A. Krassimir, "Intuitionistic fuzzy logics as tools for evaluation of Data Mining processes," 25th anniversary of Knowledge-Based Systems, vol. 80, pp. 122-130, 2015.

[56] M. Moniri, "Fuzzy and Intuitionistic Fuzzy Turing Machines," Fundamenta Informaticae, vol. 123, no. 3, pp. 305-315, 2013.

[57] C. Rahul, T. Arathi, L. S. Panicker, R. Gopikakumari, Morphology & word sense disambiguation embedded multimodal neural machine translation system between sanskrit and malayalam, Biomedical Signal Processing and Control 85 (2023) 105051.

[58] A. Bria, W. Faber and N. Leone, "Normal Form Nested Programs," Fundamenta Informaticae, vol. 96, no. 3, pp. 271-295, 2009.

[59] R. Jayashree, S. K. Murthy, K. Sunny, Keyword extraction based summarization of categorized kannada text documents, International Journal on Soft Computing 2 (4) (2011) 81.

[60] C. Gupta, A. Jain, N. Joshi, Fuzzy logic in natural language processing–a closer view, Procedia computer science 132 (2018) 1375–1384.

[61] G. Chen, T. T. Pham, N. Boustany, Introduction to fuzzy sets, fuzzy logic, and fuzzy control systems, Applied Mechanics Reviews 54 (6) (2001) B102–B103.

[62] H. Englund, H. Stockhult, S. Du Rietz, A. Nilsson, G. Wennblom, Learning-environment uncertainty and students' approaches to learning: A self-determination theory perspective, Scandinavian Journal of Educational Research (2022) 1–15.

[63] T. Chen, An innovative fuzzy and artificial neural network approach for forecasting yield under an uncertain learning environment, Journal of Ambient Intelligence and Humanized Computing 9 (2018) 1013–1025. .

[64] A. Chiche, B. Yitagesu, Part of speech tagging: a systematic review of deep learning and machine learning approaches, Journal of Big Data 9 (1) (2022) 1–25.

[65] C. Marsala, B. Bouchon-Meunier, Fuzzy data mining and management of interpretable and subjective information, Fuzzy Sets and Systems 281 (2015) 252–259.

[66] C. Marsala and B. Bouchon-Meunier, "Fuzzy data mining and management of interpretable and subjective information," Fuzzy Sets and Systems, vol. 281, no. Special Issue Celebrating the 50th Anniversary of Fuzzy Sets, p. 252–259, 2015.

[67] F. M. Zanzotto, L. Dell'Arciprete and A. Moschitti, "Efficient Graph Kernels for Textual Entailment Recognition," Fundamenta Informaticae, vol. Moschitti, no. 2-3, pp. 199-222, 2011.

[68] T.-L. Tseng, F. Jiang and Y. Kwon, "Hybrid Type II fuzzy system & datamining approach for surface finish," Journal of Computational Design and Engineering, vol. 2, no. 3, pp. 137-147, 2015.

[69] E. J. Khatib, R. Barco, A. Gómez-Andrades, P. Muñoz and I. Serrano, "Data mining for fuzzy diagnosis systems in LTE networks," Expert Systems with Applications, vol. 42, no. 21, p. 7549–7559, 2015.

[70] A. Estiri, M. Kahani, H. Ghaemi and M. Abasi,

"Improvement of An Abstractive Summarization Evaluation Tool using Lexical-Semantic Relations and Weighted Syntax Tags in Farsi Language," in 12th Iranian Conference on Intelligent Systems Higher Education Complex of Bam, Bam, 2014.

[71] A. Jacob, A. Babu and P. C. R. Raj, "TnT tagger with fuzzy rule based learning," in Signal Processing, Informatics, Communication and Energy Systems (SPICES), Kozhikode, 2015.

[72] A. R. Martinez, Part-of-speech tagging, Wiley Interdisciplinary Reviews: Computational Statistics 4 (1) (2012) 107–113.

[73] A. R. Martinez, "Part-of-speech tagging," Wiley Periodicals, Inc., vol. 4, pp. 107-113, 2012.

[74] H. Yamane and M. Hagiwara, "Oxymoron generation using an association word corpus and a large-scale N-gram corpus," Soft Computing, vol. 19, pp. 919-927, 2015.

[75] J. Hoon Kim, J. Seo and G. Chang Kim, "Estimating Membership Functions in a Fuzzy Network Model for Part-Of-Speech Tagging," Journal of Intelligent & Fuzzy Systems: Applications in Engineering and, vol. 4, no. 4, pp. 309-320, 1996.

[76] K. Atanassov, "Intuitionistic fuzzy logics as tools for evaluation of Data Mining processes," 25th anniversary of Knowledge-Based Systems, vol. 80, pp. 122-130, 2015.

[77] A. Chitra and A. Rajkumar, "Paraphrase Extraction using fuzzy hierarchical clustering," Applied Soft Computing, vol. 34, p. 426–437, 2015.

[78] T. J. Ross, Properties of membership functions, fuzzification, and defuzzification, Fuzzy logic with engineering applications (2010) 89–116.

[79] K. Gilda, S. Satarkar, Analytical overview of defuzzification methods, International Journal of Advance Research, Ideas and Innovations in Technology 6 (2) (2020) 359–365.

[80] P. M. LaCasse, W. Otieno, F. P. Maturana, A hierarchical, fuzzy inference approach to data filtration and feature prioritization in the connected manufacturing enterprise, Journal of Big Data 5 (2018) 1–31.

[81] L. Perumal, F. H. Nagi, Switching control system based on largest of maximum (lom) defuzzification-theory and application, Fuzzy Logic–Controls, Concepts, Theories andApplications, InTech, Rijeka (2012) 301–324.

[82] L. Perumal and F. H. Nagi, "Switching Control System Based on Largest of Maximum (LOM) Defuzzification – Theory and Application," in Fuzzy Logic – Controls, Concepts, Theories and Applications, Slavka Krautzeka, InTech, 2012, pp. 301-325.

[83] S. Naaz, A. Alam and R. Biswas, "Effect of different defuzzification methods in a fuzzy based load balancing application," IJCSI International Journal of Computer Science Issues, vol. 8, no. 5, pp. 261-267, 2011.

[84] H. Tzung-Pei, C. Chun-Hao, W. Yu-Lung and L. Yeong-Chyi, "A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions," Soft Computing, vol. 10, p. 1091–1101, 2006.

[85] C. D. Manning, "Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?," in CICLing'11 Proceedings of the 12th international conference on Computational linguistics and intelligent text processing., Tokyo, 2011.