International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 4 (2019), pp. 554-562

Development of Fuzzy Risk Score Assessment Framework for Sanctions Screening

Rogie P. Nino, Ruji P. Medina and Ariel M. Sison

Graduate School of Information Technology/Technological Institute of the Philippines – QC, Philippines

ORCIDs: 0000-0002-9334-3277(Rogie Nino), 0000-0003-1649-4634(Ariel Sison), 0000-0002-8204-8002(Ruji Medina)

Abstract

Analyzing Risk Scoring Assessment for sanctions screening is necessary to evaluate the risk weight rate of data elements involved during screening. Sanctions Screening is the process of reviewing sanctions lists to check if any investor in a fund is involved in fraud by matching the investor information (as Stop Descriptor) with the Sanctions List which contains the names of individuals who are known to be involved in financial crime or terrorism. This paper aims to develop a Fuzzy Risk Score Assessment Framework that will be used in building an appropriate fuzzy expert system to logically distribute the weight rate of risk scoring components with different match comparison result. The study will also present the inherent capability of Edit Distance algorithm or the Damerau-Levenshtein Distance algorithm to address many common misspellings and typos in string matching through insertion, deletion, transposition and substitution which are considered as a significant component of fuzzy possible success rating.

Keywords: Damerau- Levenshtein Distance, Fuzzy Logic, Risk Score Assessment, Sanctions Screening

I. INTRODUCTION

Sanctions screening is the process of reviewing sanctions lists to check if any investor in a Fund is involved in financing crime or terrorism [1]. The insurance industry plays an equally important role in combating financial crime and promoting international security. Life insurance policies and annuities have been used in money laundering and the financing of terrorism [2]. Different algorithms are being used in screening such as: 1) Soundex, a phonetic algorithm that indexes names by sound and matches them against Soundex encoded Stop Descriptors Soundex codes are four-character strings composed of a single letter followed by three numbers; 2) Metaphone, a phonetic algorithm that encodes input names with varying lengths by their English pronunciation. These encodings are compared to encoded Stop Descriptors. Metaphone algorithms are the basis for many popular spell checkers; 3) Double Metaphone, second generation of Metaphone phonetic algorithm; and 4) Fuzzy Logic, it is based on the Edit Distance algorithm (or Damerau-Levenshtein Distance algorithm) in which using Stop Descriptors; the algorithm can address many common misspellings and typos such as insertion, deletion, substitution and transposition.

The Damerau-Levenshtein distance algorithm is a popular method of fuzzy string matching. It is a string metric for measuring the difference between two sequences. The likelihood of a match being either "true match" or "false positive" during sanctions screening process will depend on the Risk Score Assignment (RSA).

The Risk Score Assignment (RSA) will be the basis for each match being generated during the screening process. It compares the data elements or components involved with the match and based on those elements' similarities/differences produces a highly-tunable score that represents the likelihood of a match being either "true" or "false positive". The problem is the risk weight rate being applied across sanctions screening wherein the distribution percentage may not be logically viable for different matching result of risk scoring. To mitigate the risk, this paper aims to develop a Fuzzy Risk Score Assessment Framework for sanctions screening. The new framework will also take advantage the utilization of Damerau-Levenshtein Distance algorithm which can be used to calculate the number of changes made in one text string (Stop Descriptor) to match from the Sanctions List or Watch List.

The rest of this paper is organized as follows: Section II is the discussion of related works to the proposed fuzzy logic framework; Section III presents and describes the proposed fuzzy logic framework and explains the implementation details. Experimental setup and results and implementations are also provided; lastly Section IV concludes the paper and recommended study for future works.

II. REVIEW OF RELATED LITERATURES

This section presents some related works that are relevant to the Sanctions Screening, Risk Scoring, Fuzzy Logic Implementation and Damerau-Levenshtein Distance algorithm.

A. Sanctions Screening

The Sanctions Screening (SS) is the process of reviewing sanctions lists or also referred as "watchlists". It is a compilation of multiple regulatory and enhanced due diligence lists from all major sanctioning bodies around the world, including global lists such as OFAC, UN sanctions, EU sanctions, HM Treasury and PEP, and in-country lists [3]. The sanction lists may contain different entities to ensure that there is no a breach of Country Sanctions Programs by considering the following information during sanction screening validation: 1) Extent of policy coverage; 2) Country of residence for an individual 3) Name, address, and date of birth; 4) Nationality of an individual; 5) Country of registration for a company; 6) Domicile of a company; 7) Location of loss; 8) Destination of travel; 9) Shipments to and from and transits through sanctioned countries; and 10) Recipient of goods if cargo is to be delivered. Currently, some of the companies are limiting the number entities to be checked in sanctions screening particularly to those with high volume number of policies and claim transactions being screened and its risk score assignment may vary depends on the risk score rating result of the identified entities or components. Table 1 is a sample of Non-Fuzzy Logic Risk Score Assessment assignment based on the selected entities or components to be screened in sanctions screening.

Table 1: Sample Risk Score Card Assignment

Components	Weight Rate
Full Name (FY)	50%
Birth Year (BY)	20%
Associated Country (AC)	30%

Table 2 is a sample of Full Name Comparison of Risk Score computation which compares the input record's name (Stop Descriptor) to each name associated with the matched Sanctions List or Watch List entity. The "full name comparison" will vary depends on the fuzzy logic percentage that will set-up in running the screening process wherein a 100% will denote an exact matching (verbatim). This scoring component will compare all words from both the input record marked as a "Name field" and all words in the entity name fields on sanctions list or watch list entity. When the amount of words differs between Input Record Name and Watch List Entity Name, the scoring component will always produce a score based on the element with the least number of tokens. In the below example the first two rows produce a score of 100% because all the Input Record Name's words appear in the Watch List Entity Name's words.

Table 2. Full Name Comparison Risk Score Computation

Record Name	Watch List Entity Name	Risk Score
Juan dela Cruz	Juan dela Cruz	100%
Juan dela Cruz	Juan G. dela Cruz	100%
Dr. Juan dela Cruz	Juan dela Cruz	100%
Dr. Juan dela Cruz	Juan Santos Cruz	67%

Table 3 is sample of Birth Year comparison which compares the Input Record's Birth (BY) to each BY associated with the matched Watch List entity (i.e. BY, alternate BY).

Table 3. Birth Year Comparison Risk Score Computation

Record Name	Risk Score
Exact Match	100%
+/- 1 Year	75%
+/- 2 Years	50%
Over 2 Years	0%

Associated Country comparison compares the Input Record's Associated Country (ISO Country Code or Country Name) to each country associated with the matched Watch List entity (i.e. Country, Nationality, Citizenship). The risk scores will be then multiplied in the Weight Rate for each component to get the total Risk Score Result.

Other Weight Risk Scoring Assignment using Ranges from 1 to 10

These weights are relative to the highest weight assigned to a scoring component: If the highest value for any component is "8", a weight of "4" for another component will be 50% as impactful and "2" will be 25% as impactful. If the highest value for any component is "10", a weight of "4" for another component will be 40% as impactful and "1" will be 10% as impactful.

B. Edit Distance and Damerau-Levenshtein Distance Algorithm

Edit distance is a measure of similarity between two strings evaluated based on the minimum number of operations required to transform one string into the other while **Damerau–Levenshtein** is distance between two words and the minimum number of operations *(consisting of insertions, deletions or substitutions of a single character, or transposition of two adjacent characters)* required to change one word into the other. The Damerau–Levenshtein distance differs from the classical Levenshtein distance by including transpositions among its allowable operations in addition to the three-classical single-character edit operations (insertions, deletions and substitutions). The said four operations correspond to more than 80% of all human misspellings. Each spelling mistake is a wrong, missing, extra letter, or the wrong type of the order of two different consecutive letters, for example, "ab" typed as "ba" is considered as 1 mistake while it is 2 according to Levenshtein edit-distance [4]. Figure 1 illustrates the Damerau–Levenshtein distance between two strings $\boldsymbol{a}$ and $\boldsymbol{b}$ a function $d_{a,b}(i, j)$ is defined, whose value is a distance between an $\boldsymbol{i}$–symbol prefix (initial substring) of string $\boldsymbol{a}$ and a $\boldsymbol{j}$–symbol prefix of $\boldsymbol{b}$. It is where $\boldsymbol{1}_{(ai \neq bi)}$ is the indicator function equal to 0 when $\boldsymbol{a_i} = \boldsymbol{b_j}$ and equal to **1** otherwise.

$$
d_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} d_{a,b}(i-1,j)+1 \\ d_{a,b}(i,j-1)+1 \\ d_{a,b}(i-1,j-1)+1_{(a_i \neq b_i)} \\ d_{a,b}(i-2,j-2)+1 \end{cases} & \text{If } i,j > 1 \text{ and } a_i \text{ and } b_{j-1} \text{ and } a_{i-1} \text{ and } b_j \\ \min \begin{cases} d_{a,b}(i-1,j)+1 \\ d_{a,b}(i,j-1)+1 \\ d_{a,b}(i-1,j-1)+1_{(a_i \neq b_i)} \end{cases} & \text{otherwise.} \end{cases}
$$

Fig. 1. Recursive Function Representation of Damerau-Levenshtein Distance

Each recursive call matches one of the cases covered by the Damerau–Levenshtein distance above.

$\boldsymbol{d_{a,b}\ (i-1, j) + 1}$ (1)

- corresponds to a deletion (from a to b).

$\boldsymbol{d_{a,b}\ (i, j-1) + 1}$ (2)

- corresponds to an insertion (from a to b).

$\boldsymbol{d_{a,b}\ (i-1, j-1) + 1_{(aj \neq bi)}}$ (3)

- corresponds to a match or mismatch, depending on whether the respective symbols are the same.

$\boldsymbol{d_{a,b}\ (i-2, j-2) + 1}$ (4)

- corresponds to a transposition between two successive symbols.

C. Fuzzy Logic and Related Works

While classical logic operates with only two values 1 (true) and 0 (false), fuzzy logic extended the range of truth values to all real numbers in the interval between 0 and 1. The interval of this number represents the possibility that a given statement was true or false. Fuzzy logic is determined as a set of mathematical principles for knowledge representation based on degrees of membership rather than on crisp membership of classical binary logic [5].

There are several designs and study that has been proposed with regards to fuzzy logic assessment that leads to decision making and or valuation analysis. These studies will be the basis of the researcher on the viability of applying fuzzy logic in risk scoring assessment for sanctions screening by developing Fuzzy Risk Score Assessment Framework process flow in building an appropriate fuzzy expert system. A research study presents how a mobile application using Fuzzy Logic and Global Positioning System (GPS) analyzes a student's lifestyle and provides recommendations and suggestions based on the results [6]. Study developed a new product screening evaluation framework that will be used for a Go/No Go decision at the Front End. A new product screening using fuzzy logic in which the criteria ratings and their corresponding importance are assessed in linguistic terms described by fuzzy numbers, and fuzzy weighted average is employed to aggregate these fuzzy numbers into a fuzzy-possible-success rating (FPSR) of the product [7]. Fuzzy-logic based model for the screening of Obstructive sleep apnea (OSA) model were developed for accurately assessing the risk of OSA by assigning weights to certain predictors that play a greater role in OSA [8]. Developed a Multi-criteria Decision Making (MCDM) approach to select and evaluate suppliers wherein the ratings of alternatives and important weights of criteria are expressed in linguistic terms using generalized fuzzy numbers. To make procedure easier and more practical, the weighted ratings are defuzzified into crisp values by employing the "maximizing" and "minimizing" set ranking approach to determine the ranking order of alternatives [9]. The paper conducted a comparative study which aimed to compare measurements of the consumer and the product regardless of brands' size using with fuzzy approaching manner and concluded with a fitness ratio in terms of fuzzy numbers. By this way, online shoppers will be able to find best fitted products for their body measurements in each brand [10].

III. PROPOSED RISK SCORE ASSESSMENT FUZZY

Sanctions screening is a process used to screen real-time and or batch file transactions against Watch List to determine if economic and regulatory sanctions are to be applied against a person. The basis of matching depends on the Stop Descriptor (phrase that caused the match) and Match Parameters which includes the Matched Field, Match Type (Watch List to be compared), and Match Score (or Risk Score Assignment). To assess the appropriate RSA, this paper will use the Fuzzy Expert System using the Fuzzy Risk Score Assessment Framework shown in Figure 2.

Below framework will conform if the following process of developing a fuzzy expert system is applicable.

1. Specify the problem and define linguistic variables – (Linguistic Assessment)
2. Determine fuzzy sets – (Translation of Linguistic Assessment)
3. Elicit and construct fuzzy rules - (Translation of Linguistic Assessment)
4. Encode the fuzzy sets, fuzzy rules and procedures to perform fuzzy inference into the expert system – (Fuzzy Number Aggregation and Inference)
5. Evaluate and tune the system – (Evaluate and Tune the System)

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 4 (2019), pp. 554-562

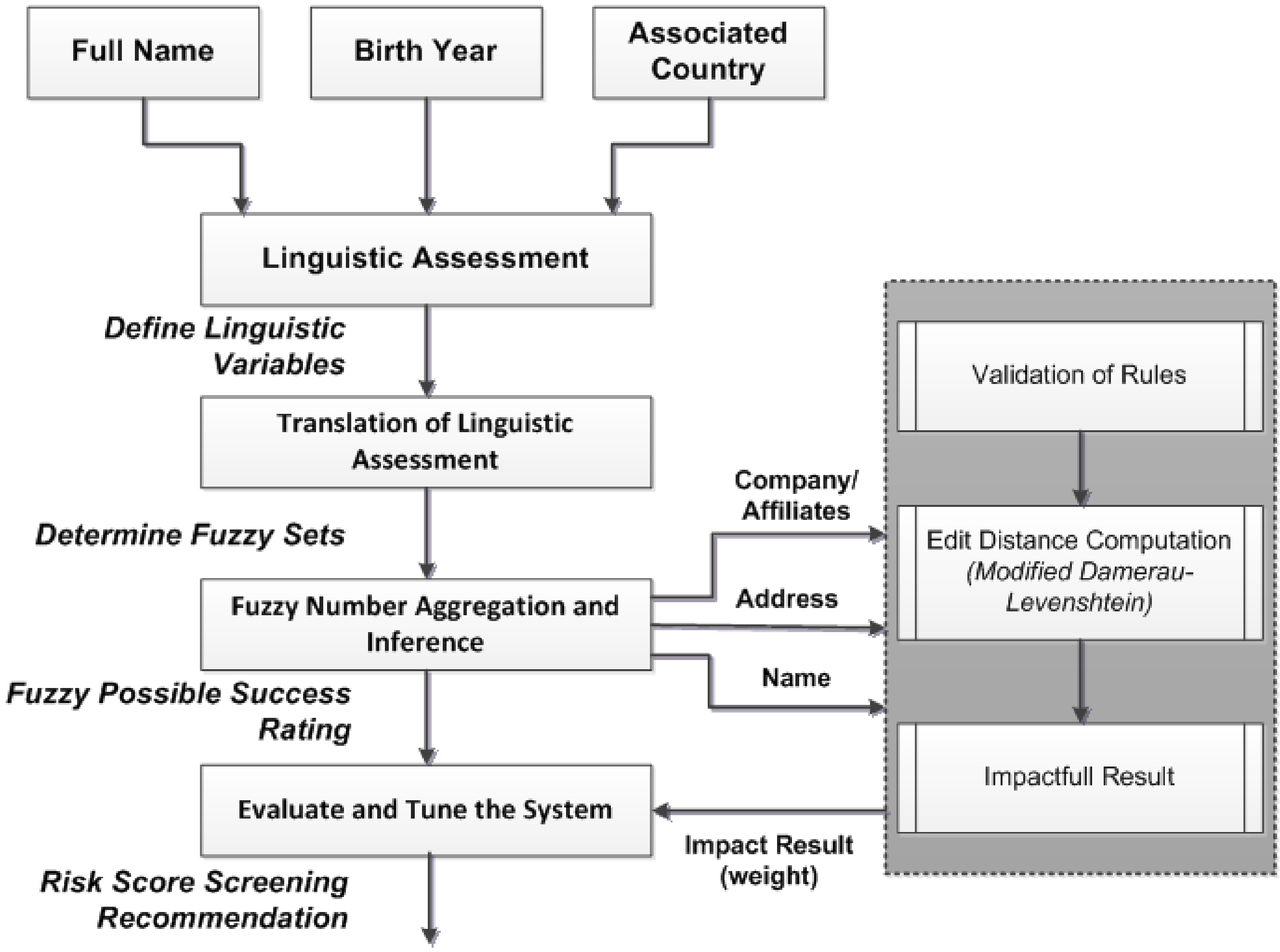

Fig. 2. Fuzzy Risk Score Assessment Framework.

The Damerau- Levenshtein Distance will be used in the process of fuzzy inference involves all of the entities that are described in Membership Functions, Logical Operations, and If-Then Rules of Risk Score Assessment Fuzzy Logic in sanctions screening. The development and testing of the application using the Damerau- Levenshtein Distance algorithm will undergo assessment process utilizing metrics or the string match (Full Name and Associated Country) weight result.

To assess the viability of implementing fuzzy logic in sanctions screening, MATLAB tool was utilized to evaluate the Fuzzy Risk Score Assessment Framework shown in Figure 2.

A. Linguistic Assessment

Fuzzy set theory lies the idea of linguistic variables. A linguistic variable is a fuzzy variable. The range of possible values of a linguistic variable represents the universe of discourse of that variable. In this study, there are three main linguistic variables in Fuzzy Risk Score Assessment:

1. Full Name comparison (FN)
2. Birth Year Comparison (BY) and
3. Associated Country Comparison (AC)

A linguistic variable carries with it the concept of fuzzy set qualifiers, called hedges. Hedges are terms that modify the shape of fuzzy sets. The linguistic values we've considered are Very Significant, Significant, Adequate, Marginal and Weak.

Table 4. Linguistic Variable and Their Ranges - Full Name (FN)

Linguistic Value	Notation	Input Consideration	Numerical Range
Significant	S	FN Comparison Percentage	[0.95,1.00]
Adequate	A	FN Comparison Percentage	[0.93,0.97]
Weak	W	FN Comparison Percentage	[0.91,0.95]

Table 5. Linguistic Variable and Their Ranges - Birth Year (BY)

Linguistic Value	Notation	Input Consideration	Numerical Range
Significant	S	BY Comparison Percentage	[0.63,1.00]
Adequate	A	BY Comparison Percentage	[0.38,0.75]
Weak	W	BY Comparison Percentage	[0.13,0.50]
Insignificant	I	BY Comparison Percentage	[0.00,0.25]

Table 6. Linguistic Variable and Their Ranges - Associated Country (AC)

Linguistic Value	Notation	Input Consideration	Numerical Range
Very Significant	VS	AC Comparison Percentage	[0.90,1.00]
Significant	S	AC Comparison Percentage	[0.85,0.95]
Adequate	A	AC Comparison Percentage	[0.80,0.90]
Weak	W	AC Comparison Percentage	[0.75,0.85]
Insignificant	I	AC Comparison Percentage	[0.70,0.80]

In insurance, Full Name can be a policy holder (pre-bind) and or beneficiary (claims) of the policy. In Table IV, this full name will be matched to the Sanction List or Watch List to calculate the degree of matching in terms of percentage wherein 100% is considered as exact match. The result may vary to the acceptable linguistic value if the comparison is significant, adequate or weak with an overlapping and interval value of 5% for 90% to 100% percentage range. In Table V, Birth Year is difference between the birthdate of a policy holder or beneficiary of the year of Birth of an entity in Sanction List or Watch List. The calculation is not only applied in the Year, but the Month and Date are also variables in the computation. The Birth Year will be tag as Significant Logistic value while the difference of +/-1 year will be Adequate, +/-2 will be Weak and over 2 years will be immaterial. In Table VI, Associated Country on other hand are the birth place, address, and address of work of the policy holder or beneficiary not limited to City, Province and Country. Same with Full Name it will be matched to the Sanction List or Watch List to calculate the degree of matching in terms of percentage wherein the linguistic values are Very Significant, Significant, Adequate, Marginal and Weak with an overlapping and interval value of 5% from 70% to 100% percentage range.

B. Translation of Linguistic Assessment

Following the theory of Lofti A. Zadeh wherein it describes a fuzzy set **A** in **X** is characterized by a membership function ***f*A(*x*)** which associates with each point in **X** a real number in the interval **[0,1]**, with the values of ***f*A(*x*)** at ***x*** representing the "grade of membership" of ***x*** in **A**., the following component value of sanctions screening represents ***f*A(*x*)** where **x** is the resulting value of percentage matching between of FN and AC to the Sanction List or Watch List and the year difference of BY. Figure 3, Figure 4, and Figure 5 are the graphical representations of the grade of membership of ***x*** in **A**.

Figure 3 shows a fuzzy number for approximate linguistic effect rating values for Full Name Matching as Significant with numerical range from 0.95 to 1.00, Adequate with numerical range from 0.93 to 0.97 and Weak with numerical range from 0.91 to 0.95. Figure 4 shows a fuzzy number for approximate linguistic effect rating values for Birth Year as Significant with numerical range from 0.63 to 1.00, Adequate with numerical range from 0.38 to 0.75, Weak with numerical range from 0.13 to 0.50 and Insignificant with numerical range from 0.00 to 0.25.

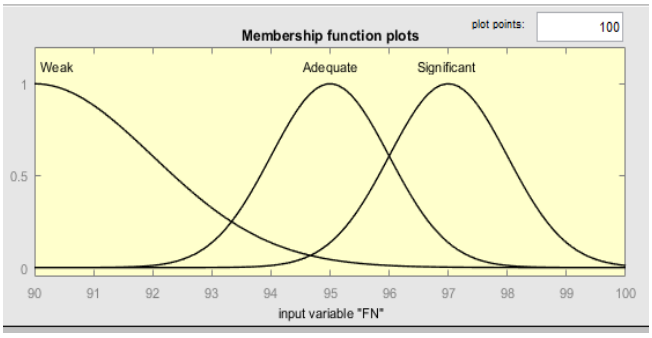

Fig. 3. Degree of Membership: Full Name Comparison.

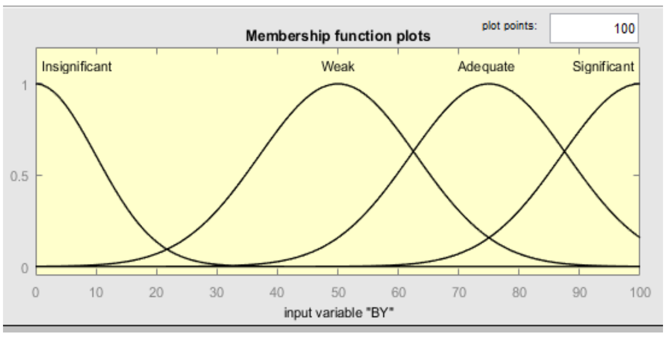

Fig. 4. Degree of Membership: Birth Year Comparison.

Figure 5 shows a fuzzy numbers for approximate linguistic effect rating values for Associated Country Matching as Very Significant with numerical range from 0.90 to 1.00, Significant with numerical range from 0.85 to 0.95, Adequate with numerical range from 0.80 to 0.90, Marginal with numerical range from 0.75 to 0.85 and Weak from 0.70 to 0.80.

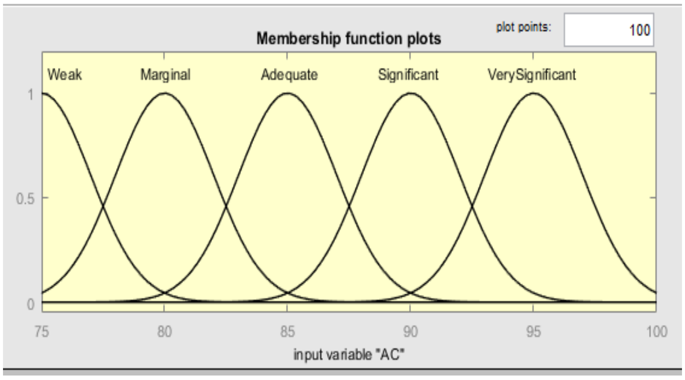

Fig. 5. Degree of Membership: Associated Country Comparison.

C. *Fuzzy Number Aggregation and Inference*

Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned. The process of fuzzy inference involves all the pieces that are described in Membership Functions, Logical Operations, and If-Then Rules. [11] Below are the rules applied in identifying the risk scoring of sanctions screening components (Full Name-FN, Birth Year-BY and Associated Country-AC) with expected total 100% percentage and distributed to the said sanctions screening components.

1. If (FN is Significant) then (RISK_SCORE is Risk_Score_5)
2. If (FN is Adequate) then (RISK_SCORE is Risk_Score_4)
3. If (FN is Weak) then (RISK_SCORE is Risk_Score_3)
4. If (BY is Significant) then (RISK_SCORE is RiskScore4)
5. If (BY is Adequate) then (RISK_SCORE is RiskScore3)
6. If (BY is Weak) then (RISK_SCORE is RiskScore2)
7. If (BY is Insignificant) then (RISK_SCORE is RiskScore1)
8. If (AC is Very Significant) then (RISK_SCORE is RiskScore5)
9. If (AC is Significant) then (RISK_SCORE is RiskScore4)
10. If (AC is Adequate) then (RISK_SCORE is RiskScore3)
11. If (AC is Marginal) then (RISK_SCORE is RiskScore2)
12. If (AC is Weak) then (RISK_SCORE is Risk Score 1)

For FN, Risk Score 5 denotes 50% risk weight as impactful during screening processing while Risk Score 4 and Risk Score 3 denotes 50% and 40% as impactful respectively.

For BY, Risk Score 4 denotes 20% risk weight as impactful during screening processing while Risk Score 3 denotes 15%, Risk Score 2 denotes 10%, and Risk Score 1 denotes 5% as impactful.

For AC, Risk Score 5 denotes 30% risk weight as impactful during screening processing while Risk Score 4 denotes 24%, Risk Score 3 denotes 18%, Risk Score 2 denotes 12% and Risk Score 1 denotes 6% as impactful.

Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set. Aggregation only occurs once for each output variable, just prior to the fifth and final step, defuzzification. The input of the aggregation process is the list of truncated output functions returned by the implication process for each rule. The output of the aggregation process is one fuzzy set for each output variable. [11] Using MATLAB tool, below are the sample aggregation result that will be considered to build the required fuzzy expert system for Fuzzy Risk Score Assessment for sanctions screening.

Figure 6 shows that FN input nearing 97% will have a Risk Score of approximately 5. Mean that FN will be assigned 50% as impactful during Sanctions Screening. Figure 7 shows that BY input nearing 75% will have a Risk Score of approximately 1.5. Mean that FN will be assigned 15% as impactful during Sanctions Screening.

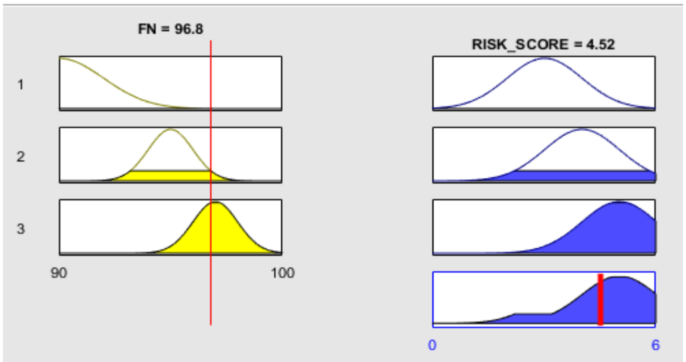

Fig. 6. Risk Score: Full Name comparison (FN).

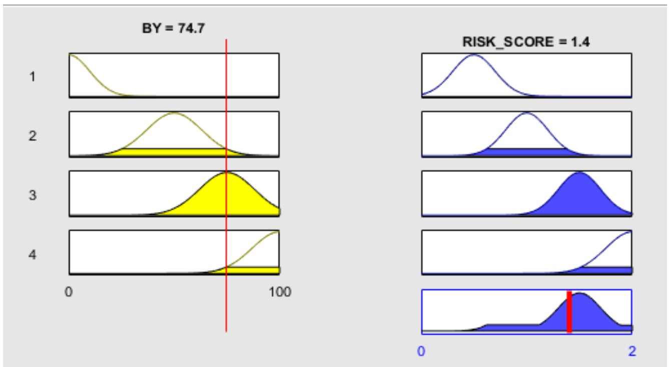

Fig. 7. Risk Score: Birth Year Comparison (BY).

Figure 8 shows that AC input nearing 90% will have a Risk Score of approximately 2.4. Mean that FN will be assigned 24% as impactful during Sanctions Screening. With this, new Fuzzy Weight Rate below in Table 7 (Fuzzy Weight Rate for FN), Table 8 (Fuzzy Weight Rate for BY) and Table 9 (Fuzzy Weight Rate for AC) will be established.

Table 7. Fuzzy Weight Rate - Full Name (FN)

Linguistic Value	Numerical Range	Approx. Weight Rate
Significant	[0.95,1.00]	50%
Adequate	[0.93,0.97]	40%
Weak	[0.91,0.95]	30%

Table 8. Fuzzy Weight Rate - Birth Year (BY)

Linguistic Value	Numerical Range	Approx. Weight Rate
Significant	[0.63,1.00]	20%
Adequate	[0.38,0.75]	15%
Weak	[0.13,0.50]	10%
Insignificant	[0.00,0.25]	5%

Table 9. Fuzzy Weight Rate - Associated Country (AC)

Linguistic Value	Numerical Range	Approx. Weight Rate
Very Significant	[0.90,1.00]	30%
Significant	[0.85,0.95]	24%
Adequate	[0.80,0.90]	18%
Weak	[0.75,0.85]	12%
Insignificant	[0.70,0.80]	6%

D. Evaluate and Tune the System

Tuning is the most laborious and tedious part in building a fuzzy system. It often involves adjusting existing fuzzy sets and fuzzy rules. As this is true in the development of expert system, embedding the principle of fuzzy in Risk Score Assessment and the likelihood of a matching process in sanctions screening being either the result is “true match” or “false positive” will provide significant impact and result versus to the constant risk weight rate combination used across all entities during sanctions screening.

The sample risk weight rate combination of 50% Full Name (FN), 30% Associated Countries (AC) and 20% Birth Year (BY) will vary depends on the data input identified in linguistic variable and ranges.

Result Comparison 1, 2 and 3 in Table 10, 11 and 12 respectively shows that applying fuzzy logic principle in Risk Score Assessment in sanctions screening will provide significant value *(Non-Fuzzy Risk Rate Result < Fuzzy Risk Rate Result)* in determining the matching combination probability of being either the result is “true match” or “false positive” during the sanctions screening process.

Table 10. Result Comparison 1

Component	Result of Matching	Non-Fuzzy Risk Weight Rate	Non-Fuzzy Risk Rate Result	Fuzzy Risk Weight Rate	Fuzzy Risk Rate Result
FN	99.00%	50%	49.50%	50%	49.00%
BY	50.00%	20%	10.00%	10%	5.00%
AC	95.00%	30%	28.50%	30%	28.50%
Result of Likelihood of “True Match””		100%	88.00%	90%	92.22%

Table 11. Result Comparison 2

Component	Result of Matching	Non-Fuzzy Risk Weight Rate	Non-Fuzzy Risk Rate Result	Fuzzy Risk Weight Rate	Fuzzy Risk Rate Result
FN	90.00%	50%	45.00%	40%	36.00%
BY	75.00%	20%	15.00%	15%	11.25%
AC	90.00%	30%	27.00%	24%	21.60%
Result of Likelihood of “True Match””		100%	87.00%	79%	87.15%

Table 12. Result Comparison 3

Component	Result of Matching	Non-Fuzzy Risk Weight Rate	Non-Fuzzy Risk Rate Result	Fuzzy Risk Weight Rate	Fuzzy Risk Rate Result
FN	99.00%	50%	49.50%	50%	49.50%
BY	75.00%	20%	15.00%	15%	11.25%
AC	100.00%	30%	30.00%	30%	30.00%
Result of Likelihood of “True Match””		100%	94.50%	95%	95.53%

The Damerau-Levenstein Distance algorithm was implemented in assessing the likelihood of Full Name and Associated Country entities.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 4 (2019), pp. 554-562

Input

String 1: **Rogie Plmaro Nion**

String 2: **Rogie Palmarion Nino**

Process

Begin

Step 1: *Check for empty string*

$m = | s |$

$n = | t |$

if $m = 0$ *return n end-if*

if $n = 0$ *return n end-if*

Step 2: *Create distance matrix d*

for $i = 0$ *to m:* $d_{i,0} = i$ *end-for*

for $j = 0$ *to n:* $d_{0,j} = j$ *end-for*

Step 3: *Calculate distance matrix d*

for $i = 1$ *to m*

for $j = 1$ *to n*

if $s_{i-1} = t_{j-1}$

$d_{i,j} = d_{i-1,j-1}$

else

$d_{i,j} = \min (d_{i-1,j}, d_{i,j-1}, d_{i-1,j-1}) +1$

if $i>1$ *and* $j>1$

if $s_{i-1} = t_{j-2}$ *and* $s_{i-2} = t_{j-1}$

$d_{i,j} = \min (d_{i,j}, d_{i-2,j-2}, +1)$

end-if

end-if

end-if

end-for

end-for

return dm,n

End

Figure 9 shows the Damerau-Levenshtein matrix. The intersections in light blue of character columns and rows in the white area are the optimal calculated distances for all substrings. With the above input strings, the matrix will give a Damerau- Levenshtein Distance Result of 4 because the intersection of "Rogie Plmaro Nion"and "Rogie Palmario Nino" is 4.

		R	o	g	i	e		P	a	l	m	a	r	i	o		N	i	n	o
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
R	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
o	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
g	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
i	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
e	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13
P	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12
l	8	7	6	5	4	3	2	1	2	1	2	3	4	5	6	7	8	9	10	11
m	9	8	7	6	5	4	3	2	3	2	1	2	3	4	5	6	7	8	9	10
a	10	9	8	7	6	5	4	3	2	3	2	1	2	3	4	5	6	7	8	9
r	11	10	9	8	7	6	5	4	3	4	3	2	1	2	3	4	5	6	7	8
o	12	11	10	9	8	7	6	5	4	5	4	3	2	3	2	3	4	5	6	7
	13	12	11	10	9	8	7	6	5	6	5	4	3	4	3	2	3	4	5	6
N	14	13	12	11	10	9	8	7	6	7	6	5	4	5	4	3	2	3	4	5
i	15	14	13	12	11	10	9	8	7	8	7	6	5	4	5	4	3	2	3	4
o	16	15	14	13	12	11	10	9	8	9	8	7	6	5	4	5	4	3	4	3
n	17	16	15	14	13	12	11	10	9	10	9	8	7	6	5	6	5	4	3	4

Figure 9. Damerau- Levenshtein Distance Result Matrix

To further explain, for example, the distance between "Rogie Palmari" and "Rogie Plmar" is 2 because the intersection at **second** 'i' and 'r' respectively is 2, which is the number of edits necessary to convert one to the other.

To calculate the result of matching percentage, the computed Distance was used in the formula below.

Fuzzy Weight Rate = 100- (Distance/Length of String2) (5)

The new fuzzy weight rate will be used to evaluate the risk rate of data elements involved in a match to assess the similarities/differences which produces a highly-tunable score that represents the likelihood of a match being either "true" or "false positive".

Above simulations were also utilized in computing the Fuzzy Weight Rate of Associated Country entity.

V. CONCLUSION

With the use of fuzzy logic principle on the Risk Score Assessment of sanctions screening process to determine the likelihood of a match between the policy holder and claimant data against Sanctions List or Watch List being either "true match" or "false positive" shows significant increase with regards to its probability of capturing transactions sanctioned (e.g. country, entity, organization or individual).

For future study, it is recommended to consider of making robust Risk Score Assessment rating which learn the fuzzy risk weight rate on its own. The machine learning algorithms such as support vector machines and neural networks can be considered. Whilst the Damerau-Levenshtein Distance is one of the available algorithms to be used in string matching, the potential of optimizing its current algorithm will help the performance of screening process specifically involving two longer strings.

REFERENCES

[1] "IPES," 2017. [Online]. Available: https:// https://www.ipes.com/sites/default/files/uploads/Sanctions%20Screening%20Leaflet_March%2015.pdf

[2] "PWC," 2017. [Online]. Available: https://www.pwc.com/us/en/financial-services/publications/assets/global-sanctions-practice-insurance.pdf

[3] "Trulioo," 30 August 2016. [Online]. Available: https://www.trulioo.com/blog/aml-watchlist-screening

[4] Quyet Thang Dang, Trung Huy Phan, "Determining Restricted Damerau-Levenshtein Edit-Distance of Two Languages by Extended Automata", *2010 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future (RIVF),* 2010

[5] M. Negnevitsky, Artificial Intelligence: *A Guide to Intelligent Systems, 2nd edition, 2005*

[6] S. Ghosh, A. S. Boob, N. Nikhil, N. R. Vysyaraju, and A Kumar, "A Fuzzy Logic System to Analyze a Student's Lifestyle", in *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI),* Feb 4-6, 2017, Pages 231 - 236, Doha Qatar

[7] C. T. Lin, and C. T. Chen," A Fuzzy-Logic-Based Approach for New Product Go/NoGo Decision at the Front End", *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, January 2004, IEEE Journals & Magazines, Volume 34, Issue 1, Pages 132 - 142*

[8] J. M. Matthews, M. Kwiatkowska, and L. R. Matthews, "Preliminary Fuzzy Model for Screening Obstructive Sleep Apnea*", IEEE, 2013, Pages 187 - 191*

[9] L.H. Van, V.F. Yu, S. Chou, and L.Q. Dat, "Supplier Selection and Evaluation Using Generalized Fuzzy Multi-Criteria Decision-Making Approach", *2016 Eighth International Conference on Knowledge and Systems Engineering (KSE), 2016, Pages 31 - 36*

[10] E. Nasibov, M. Demir, B. Okur, and A. Vahaplar, "A Fuzzy Logic Approach to Predict the Best Fitted Apparel Size in Online Marketing", *Computer Science and Engineering (UBMK) 2017 International Conference on, pp. 38-42, 2017*

[11] "Mathworks," 2017. [Online]. Available: https://www.mathworks.com/help/fuzzy/fuzzy-inference-process.html