Notes on Intuitionistic Fuzzy Sets
ISSN 1310–4926
Vol. 21, 2015, No. 3, 48–53

Similarity and dissimilarity of whole genomes using intuitionistic fuzzy logic

Subhram Das[1], Debanjan De[2], D. K. Bhattacharya[3]

[1] Computer Science & Engineering Department, Narula Institute of Technology
Kolkata 700109, India
E-mail: subhram@gmail.com

[2] Quality Control Officer, Pest Control India
Kolkata, India
E-mail: debanjande85@gmail.com

[2] Emeritus Professor, Rabindra Bharati University
Kolkata, India
E-mail: dkb_math@yahoo.com

Abstract: Whole genomes in general and poly-nucleotides, in particular, have an interesting representation in twelve dimensional hypercube I^{12} based on fuzzy set theory, but it has some limitations and drawbacks. With a view to removing such drawbacks of the representation, the present paper derives some new representation of whole genomes and poly-nucleotides based on Intuitionistic Fuzzy set theory and shows that that such a representation is free from any such limitation as mentioned above. Finally it applies the new representation in testing similarity/ dissimilarities of whole genomes and polynucleotides.

Keywords: Intuitionistic Fuzzy set, Intuitionistic Fuzzy Polynucleotide space, Similarity/ Dissimilarities of whole genomes and poly-nucleotides, Metric.

AMS Classification: 03E72.

1 Introduction

Necessity of introducing fuzzy set theory is realized in the process of representing a polynucleotide consisting of finite number of codons on a single hypercube I^{12}. This is the background of fuzzy polynucleotide space as introduced by Torres and Nieto (2003) [1]. Torres and Nieto (2003) [3] introduced the notion of fuzzy polynucleotide space based on the principle of the fuzzy hypercube of Kosko, (1992) [9]. The idea of differentiating polynucleotide and whole genomes on the basis of fuzzy set theory is well understood from the

work of Angela Torres and Juan J. Nietoin (2003) [1], where they used the metric as introduced in (2000) [8]. With the help of this metric they could differentiate polynucleotides and some whole genomes. Later on, in (2006) [2] different types of metric were used for comparison of polynucleotides and whole genomes. They could show that in all cases the metrics behaved similarly. In [10] the present authors cited some examples of whole genomes, where all the metrics mentioned above did not behave similarly. The possible reason is that whenever we understand the frequencies of polynucleotide and whole genome on the unit 12 dimensional hypercube, the information is not complete, as it does not consider hesitation factor, which is always present in real situation. In fact Intuitionistic Fuzzy set (IFS) concept is more robust than that of Fuzzy set; because it always accommodates some kind of hesitancies. This is the motivation in using Intuitionistic Fuzzy Set theory in place of Fuzzy Set theory in comparing sequences of whole genomes.

For the sake of comparison of our new methods using Intuitionistic fuzzy set with the earlier methods of [10] involving fuzzy set theory, we take the same polynucleotides S1, S2, S3 and the same four whole genomes as given in [10].

2 Preliminaries

Intuitionistic Fuzzy Sets [4, 5] are generalization of Fuzzy sets [6] in which non-membership values are not obtainable from the membership values, rather both of them have to be specified separately**.**

Let X be a non-empty set. An Intuitionistic fuzzy set A on X is defined as $A = \{< x, \mu_A(x), \nu_A(x) > | \, x \in X\}$, where the functions $\mu_A : X \to [0,1]$ and $\nu_A : X \to [0,1]$ define respectively the degree of membership and the degree of non-membership of the element x in X to the set A, and $0 \le \mu_A(X) + \nu_A(X) \le 1$ for each x in X. Obviously an ordinary fuzzy set can be written as $\{< x, \mu_A(x), 1 - \mu_A(x) > | \, x \in X\}$,

In reality non-membership is always associated with some sort of hesitancy.

If we fix a fraction θ of membership value as the value of hesitancy, then it is given by $\nu_A(X) = \theta\mu_A(X)$; so non-membership value equals to $\pi_A(X) = 1 - (1+\theta)\mu_A(X)$. Hence, an intuitionistic fuzzy set can be written as $\{< x, \mu_A(x), \nu_A(x), \pi_A(x) > | \, x \in X\}$, where $\nu_A(X) = \theta\mu_A(X)$, $\pi_A(X) = 1 - (1+\theta)\mu_A(X)$.

2.1 Distance measure on Intuitionistic fuzzy set

The normalized hamming distance D_{IFS} proposed for IFS by Szmidt and Kacprzyk [7] is given by

$$D_{IFS}(A, B) = \sum_{i=1}^{n} (| \, \mu_A(x_i) - \mu_B(x_i) \, | + | \, \nu_A(x_i) - \nu_B(x_i) \, | + | \, \pi_A(x_i) - \pi_B(x_i) \, |)$$

where A and B are two IFS in X= $\{x_1, x_2, ..., x_n\}$. Obviously the general form of distance measure would be

$$D^{\alpha}_{IFS}(A, B) = \left(\sum_{i=1}^{n} (|\mu_A(x_i) - \mu_B(x_i)|^{\alpha} + |\nu_A(x_i) - \nu_B(x_i)|^{\alpha} + |\pi_A(x_i) - \pi_B(x_i)|^{\alpha}) \right)^{\frac{1}{\alpha}},$$

α is a natural number.

2.2 Similarity measures on Intuitionistic fuzzy sets [7]

$$S(A, B) = 1 - \left(1/2n \sum_{i=1}^{n} (|\mu_A(x_j) - \mu_B(x_j)|^{\alpha} + |\nu_A(x_j) - \nu_B(x_j)|^{\alpha} + |\pi_A(x_j) - \pi_B(x_j)|^{\alpha}) \right)^{\frac{1}{\alpha}}, \alpha > 0$$

2.3 Formula of Intuitionistic fuzzy representation of polynucleotide on a triplet of I^{12}

Suppose fractions of nucleotide at a point on I^{12} be given by (x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4, z_1, z_2, z_3, z_4). Then the Intuitionistic fuzzy representation of the polynucleotide A is $\{< x, \mu_A(x), \nu_A(x), \pi_A(x) > | x \in X\}$, where $\mu_A(x) = (x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4, z_1, z_2, z_3, z_4)$ $\nu_A(x) = (\theta x_1, \theta x_2, \theta x_3, \theta x_4, \theta y_1, \theta y_2, \theta y_3, \theta y_4, \theta z_1, \theta z_2, \theta z_3, \theta z_4)$, $\pi_A(x) = [\{1-(1+\theta)x_1\}, \{1-(1+\theta)x_2\}, \{1-(1+\theta)x_3\}, \{1-(1+\theta)x_4\}, \{1-(1+\theta)y_1\}, \{1-(1+\theta)y_2\}, \{1-(1+\theta)y_3\}, \{1-(1+\theta)y_4\}, \{1-(1+\theta)z_1\}, \{1-(1+\theta)z_2\}, \{1-(1+\theta)z_3\}, \{1-(1+\theta)z_4)\}]$.

2.4 Difference and similarity of polynucleotides & whole genome, using Intuitionistic fuzzy representation

Tables 1, 2 and 3 are describing Intuitionistic Fuzzy representations of polynucleotides S1, S2 and S3 respectively. Similarly Table 6 describes Intuitionistic Fuzzy representation of whole genome (a), (b), (c) & (d) [8] . Tables 4 and 5 describe the distance measure and similarity measure of polynucleotides S1, S2 and S3. Similarly Tables 7 and 8 describe distance measure and similarity measure of Intuitionistic Fuzzy Representation of Whole Genome (a), (b), (c) & (d). For simplification of calculation we take $\theta = 0.1$.

S1	1	0	0	0	0	0	.5	.5	.5	.5	0	0
	0	0	0	0	0	0	.05	.05	.05	.05	0	0
	0	1	1	1	1	1	.45	.45	.45	.45	1	1

Table 1: Intuitionistic fuzzy representation of polynucleotides S1

S2	.5	.5	0	0	0	0	.5	.5	.5	.5	0	0
	.05	.05	0	0	0	0	.05	.05	.05	.05	0	0
	.45	.45	1	1	1	1	.45	.45	.45	.45	1	1

Table 2: Intuitionistic fuzzy representation of polynucleotides S2

S3	.5	.5	0	0	.5	0	0	.5	.5	.5	0	0
	.05	.05	0	0	.05	0	0	.05	.05	.05	0	0
	.45	.45	1	1	.45	1	1	.45	,45	.45	1	1

Table 3: Intuitionistic fuzzy representation of polynucleotides S3

Distance Between	α=1	α=2	α=3	α=4	α=5	α=6	α=7	α=8	α=9	α=10
S1 & S2	2.1	1.01	0.50775	0.257525	0.131282	0.067234	0.034586	0.017867	0.009268	0.004827
S1 & S3	4.3	2.12	1.09075	0.56555	0.294439	0.153846	0.08066	0.042427	0.022385	0.011846
S2 & S3	2.2	1.11	0.583	0.308025	0.163158	0.086611	0.046074	0.024559	0.013117	0.007019

Table 4: Distance measure of intuitionistic fuzzy representation of polynucleotides S1, S2 & S3

Similarity Between	α=1	α=2	α=3	α=4	α=5	α=6	α=7	α=8	α=9	α=10
S1 & S2	0.825	0.916251	0.933518	0.940636	0.944479	0.946861	0.948466	0.949612	0.950463	0.951114
S1 & S3	0.641667	0.878665	0.914218	0.927734	0.934744	0.939	0.94184	0.94386	0.945364	0.946522
S2 & S3	0.816667	0.912203	0.930384	0.937918	0.942012	0.94457	0.946311	0.947568	0.948514	0.949249

Table 5: similarity measure of intuitionistic fuzzy representation of polynucleotides S1, S2 & S3

(a)	0.233	0.267	0.233	0.267	0.233	0.265	0.233	0.269	0.232	0.27	0.232	0.266
	0.0233	0.0267	0.0233	0.0267	0.0233	0.0265	0.0233	0.0269	0.0232	0.027	0.0232	0.0266
	0.7437	0.7063	0.7437	0.7063	0.7437	0.7085	0.7437	0.7041	0.7448	0.703	0.7448	0.7074
(b)	0.311	0.189	0.31	0.19	0.31	0.191	0.308	0.191	0.307	0.192	0.309	0.192
	0.0311	0.0189	0.031	0.019	0.031	0.0191	0.0308	0.0191	0.0307	0.0192	0.0309	0.0192
	0.6579	0.7921	0.659	0.791	0.659	0.7899	0.6612	0.7899	0.6623	0.7888	0.6601	0.7888
(c)	0.164	0.338	0.162	0.336	0.159	0.341	0.161	0.339	0.158	0.341	0.158	0.343
	0.0164	0.0338	0.0162	0.0336	0.0159	0.0341	0.0161	0.0339	0.0158	0.0341	0.0158	0.0343
	0.8196	0.6282	0.8218	0.6304	0.8251	0.6249	0.8229	0.6271	0.8262	0.6249	0.8262	0.6227
(d)	0.248	0.248	0.228	0.276	0.249	0.248	0.225	0.278	0.246	0.253	0.224	0.277
	0.0248	0.0248	0.0228	0.0276	0.0249	0.0248	0.0225	0.0278	0.0246	0.0253	0.0224	0.0277
	0.7272	0.7272	0.7492	0.6964	0.7261	0.7272	0.7525	0.6942	0.7294	0.7217	0.7536	0.6953

Table 6: Intuitionistic fuzzy representation of whole genome (a),(b),(c) & (d)

Distance Between	α=1	α=2	α=3	α=4	α=5	α=6	α=7	α=8	α=9	α=10
(a) & (b)	2.0196	0.155964	0.012543	0.001015	8.24E-05	6.7E-06	5.47E-07	4.47E-08	3.66E-09	3.01E-10
(a) & (c)	1.9096	0.139554	0.01063	0.000815	6.28E-05	4.85E-06	3.76E-07	2.93E-08	2.28E-09	1.79E-10
(a) & (d)	0.3256	0.004555	7.19E-05	1.2E-06	2.09E-08	3.72E-10	6.76E-12	1.24E-13	2.32E-15	4.37E-17
(b) & (c)	3.9292	0.590183	0.092292	0.01452	0.00229	0.000362	5.74E-05	9.11E-06	1.45E-06	2.31E-07
(b) & (d)	1.914	0.144056	0.011584	0.000959	8.11E-05	6.98E-06	6.09E-07	5.38E-08	4.79E-09	4.3E-10
(c) & (d)	2.0152	0.159569	0.013492	0.001174	0.000104	9.44E-06	8.66E-07	8.05E-08	7.55E-09	7.13E-10

Table 7: Distance measure of intuitionistic fuzzy representation of whole genome (a),(b),(c) & (d)

Similarity Between	α=1	α=2	α=3	α=4	α=5	α=6	α=7	α=8	α=9	α=10
(a) & (b)	0.8317	0.9671	0.9806	0.9851	0.9873	0.98856	0.98938	0.989951	0.990373	0.990696
(a) & (c)	0.8409	0.9689	0.9817	0.9859	0.9880	0.98916	0.98993	0.99047	0.990866	0.991168
(a) & (d)	0.9729	0.9944	0.9965	0.9972	0.9976	0.99777	0.99789	0.997969	0.998029	0.998073
(b) & (c)	0.6726	0.9360	0.9623	0.9711	0.9753	0.97775	0.97935	0.980467	0.981289	0.981919
(b) & (d)	0.8405	0.9684	0.9811	0.9853	0.9873	0.98848	0.98921	0.989717	0.990082	0.990358
(c) & (d)	0.8321	0.9667	0.9802	0.9846	0.9867	0.98789	0.98866	0.989185	0.989568	0.989857

Table 8: Similarity measure of intuitionistic fuzzy representation of whole genome (a),(b),(c) & (d)

3 Intuitionistic fuzzy regular weakly generalized connected spaces

Distance measures and similarity measures for different values of α show uniform results for polynucleotides and whole genomes. They also do work satisfactorily as is evidenced from the results of S_1, S_2 and S_1, S_3. As supported biologically distance between the first pair should be less than the next pair and consequently the similarity of the first pair should be greater than the next pair. Actually this has happened in our case for each value of α. As α increases, distance measures increase and similarity measures decrease. This suggests that better is the result, larger is the value of α taken.

4 Conclusions

The same four (a), (b), (c) and (d) genomes as in [10] are chosen in this paper. This is only to show that the anomalies do not occur any more if Intuitionistic Fuzzy representation is used in place of Fuzzy representation of the genome sequences. Obviously the conclusion is not true in general. The result has to be verified on a larger number of genomes in order to claim that the conclusion is general. But as some value of the parameter θ (hesitancy factor) is always

involved in the calculations, so if some contradictory result appears at all, it is only apparent. It can be adjusted by choice of suitable θ.Thus it can be definitely concluded that the Intuitionistic Fuzzy Set is one of the best tools in analyzing similarity/dissimilarities of complete genomes.

References

[1] Nieto, J. J., Torres, A., & V-T, M.M. (2003) A metric space to study differences between polynucleotides. *Appl. Math. Lett.*, 27, 1289–1294.

[2] Nieto, J.J., Torres, A., Georgiou, D.N., & Karakasidis, T.E (2006) Fuzzy Polynucleotide Spaces and Metrics. *Bulletin of Mathematical Biology*, 68, 703–725.

[3] Torres, A., & Nieto, J.J. (2003) The fuzzy polynucleotide space: Basic properties. *Bioinformatics*, 19(5), 587–592.

[4] Atanassov, K. (1986) Intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 20, 87–96.

[5] Atanassov, K. (1989) More on intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 33, 37–46.

[6] Zadeh, L.A. (1965) Fuzzy sets, *Inform. and Control* 8, 338–353.

[7] Szmidt, E & Kacprzyk, J. (2000) Intuitionistic distances between intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 114, 505–518.

[8] Sadegh-Zadeh, K. (2000) Fuzzy genomes. *Artif. Intell. Med.*, 18, 1–28.

[9] Kosko, B. (1992) *Neural Networks and Fuzzy Systems*. Prentice-Hall, Englewood Cliffs, NJ.

[10] Das, S., De, D., Dey, A. & Bhattacharya, D. (2013) Some anomalies in the analysis of whole genome sequence on the basis of Fuzzy set theory, *IJAINN*, 38–41.