Fuzzy Set Theory and Philosophical Foundations of Medicine

Julia Limberg
Medical University of Vienna
Core Unit for Medical Statistics and Informatics
Spitalgasse 23, A - 1090 Vienna, Austria
e0107842@student.tuwien.ac.at

Rudolf Seising
Medical University of Vienna
Core Unit for Medical Statistics and Informatics
Spitalgasse 23, A - 1090 Vienna, Austria
Rudolf.Seising@meduniwien.ac.at

Abstract

Dealing with notions of health, illness and disease contains dealing with fuzziness. As the paper will demonstrate, states of these notions do not only exist or not exist. The medical philosopher and physician Sadegh-Zadeh introduced the notions of fuzzy health, fuzzy illness and fuzzy disease. A closer look will be taken on the concept of fuzzy disease. Because there are different possibilities to interpret the concept of disease, amongst others by linguistic and social backgrounds, Sadegh Zadeh indroduced potential candidates: complex „human conditions“. Afterwards, fuzzifications of life sciences will be extended and the fuzzification of the genome, including two approaches that deal with this theme, will be discussed.

Keywords: Fuzzy, Disease, Gene, Genomes, DNA.

1 Introduction

Health, Illness and Disease are notions originated in the theory of medicine that can't be defined in classical logic. Therefore, Kazem Sadegh-Zadeh, a doctor and philosopher in medicine, has been discussing the meaning of these notions since the 1980s in a new way. To illustrate his ideas, he chose the fuzzy-theoretical way. ([1], p. 607):

- "health is a matter of degree,
- illness is a matter of degree,
- and disease is a matter of degree."

In 1982, within the framework of a conference of medicine and philosophy, he blurred the notion of patienthood (being afflicted by a malady) as a new notion in the theory of medicine "of which the notion of health will be the additive inverse in the following sense:

Health = 1 – patienthood" [2].

Figure 1: Kazem Sadegh-Zadeh

In 2000 Sadegh-Zadeh presented a concept based on the Fuzzy Set Theory. In his article "Fuzzy Health, Illness and Disease" [1] he fuzzified the notions "Health", "Illness" and "Disease" and in doing so, he expanded their meanings. Particularly, he paid attention to the notion of fuzzy disease, of which the interpretation and definition proves to be very complex.

In this paper, after a short summary about this medical theoretical concept of fuzzy disease another approach of Sadegh-Zadeh, "Fuzzy genomes" [14], will be adduced. In this latter approach, the literally notions will be taken slightly into the background and the interpretation of diseases through their genomes will give the structural definition of disease. Because the definition of genes or genomes is not obvious, some basics will be described. Finally, another approach about the fuzzy concept of genes that bases in many aspects on the theory of Kazem Sadegh-Zadeh will be presented shortly.

2 Disease and Fuzzy Disease

Apparently, humans have thought about the existence of disease since humans are thinking. Many points in the bible refer to diseases that are linked with unbelief and punishment.

According to Karl Eduard Rothschuh (1908-1984), the well-known German physician and historian of medicine, this metaphysical interpretation of disease is still in the human's consciousness. Amongst this view, Rothschuh also mentioned a model of disease in association with philosophy. [3] Nevertheless, this view remained much generalized and interpreting the disease as a phenomenon or a "matter of evil" can't be considered satisfactory. So, Rothschuh introduced his third model of disease: The disease as a naturalistic model.

Thus, disease can be described in two ways. Firstly, disease is treated as a real entity. In compliance with this view, disease has an autonomous existence and has its roots in something like a seed. The disease grows like plants grow and this kind of growing can be regarded as the clinical tenor of a disease. The second naturalistic view regards the disease as a consequence of a disorganization of the organism and the organism's functional and structural components.

Similar to these models, respectively interpretations, Sadegh-Zadeh combined the notions of disease. By observing linguistic and social backgrounds, he introduced potential candidates for disease – complex "human conditions" like heart attack, apoplexia, cancer, etc.

A declaration like "heart attack is a disease" is well known in common language use. Other conditions, known in society as diseases, are now taken to approximate the notion of disease. These conditions do not merely arise from the biological state of the body. They may, as well, be described as large fuzzy sets which contain many different aspects of the sick person's environment, including religion and society.

Conditions which can be described as "pain", "distress" or "feeling of loneliness" may also be considered as aspects.

To define a definition as a whole one has to take a set of human conditions (*D*) with its corresponding criteria (*C*) into account. Following rules are established:

- Every element that is member of the basic set $\{D_1, D_2, \ldots, D_n\}$ is a disease and
- Every element that is similar to a disease with respect to the criteria $\{C_1, C_2, \ldots, C_n\}$ is a disease.

The first declaration seems to be clear. However, the second declaration poses a problem, because *similarity* has to be described. For this reason, the fuzzy set difference of two fuzzy sets A and B – *differ(A, B)* – is introduced as a starting point, which is calculated as follows:

$$differ\ (A,B) = \frac{\sum_i \max(0, \mu_A(x_i) - \mu_B(x_i)) + \sum_i \max(0, \mu_B(x_i) - \mu_A(x_i))}{c(A \cup B)}$$

C, given in the denominator, is the sum of the membership values of the corresponding fuzzy set (*fuzzy set count*). For instance, there is a fuzzy set *X* with $X = \{(x, 0.6), (y, 0.9)\}$, $c(X)$ will be calculated as: 0.6 + 0.9 = 1.5.

Back to fuzzy difference: Let's assume that there is a fuzzy set *Y* with $Y = \{(x, 0.7), (y, 0.4)\}$. Fuzzy-difference *differ(X, Y)* is calculated as

$$\frac{(0+0.5)+0.1+0)}{1.6} = 0.375$$

X differs from *Y* to a degree of 0.375.

Similarity of two fuzzy sets is resulted from the inversion of fuzzy difference. According to the example above, similarity would be given as: 1 – 0.375 = 0.625.

In order to avoid comparing apples and oranges, descriptions of similarity should be reduced to assimilable subsets. For example, one raises the question how similar are the two diseases D_i and D_j considering a few criteria $\{C_1, C_2, \ldots C_m\}$.

Let's assume *A* to be a fuzzy set of arbitrary dimension and *X* as a part of this set; so *A\X*. Human conditions, like heart attack and stomach ulcer, can be arranged according to their assimilable criteria $\{C_1, C_2, \ldots C_m\}$:

- heart_ attack\{$(C_1, a_1), (C_2, a_2), \ldots, (C_m, a_m)$}
- stomach_ulcer\{$(C_1, b_1), (C_2, b_2), \ldots, (C_m, b_m)$}
- heart_ attack\{bodily_lesion, 1), (pain, 0.7), (distress, 0.8)}
- stomach_ulcer\{(bodily_lesion, 1), (pain, 0.3), (distress, 0.5)}

To calculate similarities between fuzzy sets, the following theorem is used:

$$Theorem: similar(A, B) = \frac{c(A \cap B)}{c(A \cup B)}$$

Similar comparisons include several degrees of partial (p) similarity, symbolized as p-similar($A\backslash X$, $B\backslash Y$), under the terms of the following definition:

p-similar ($A\backslash X$, $B\backslash Y$) = r, if similar (X, Y) = r.

According to the example above and using the theorem above, this would mean:

p-similar (heart_ attack$\backslash X$, stomach_ulcer $\backslash Y$) = 0.72

Assuming that $\{D_1, ..., D_n\}$ would be a small set of human conditions, because of a set of criteria $\{C_1, \ldots, C_n\}$ which these conditions have in common. Each of these conditions is interpreted in a certain society as a disease.

For this society there is an agreement of degree ε of partial similarity. This degree is a pillar of this society's notion of disease:

- Every element in the basic set $\{D_1, \ldots, D_n\}$ is a disease
- A human condition H\X is a disease, if there is a disease $D_i\backslash Y \in \{D_1, \ldots, D_n\}$ and there is a $\varepsilon > 0$, so that p-similar ($H\backslash X$, $D_i\backslash Y$) $\geq \varepsilon$

Granted, that there is the criteria set heart_attack$\backslash\{(C_1, 1), (C_2, 0.7), (C_3, 0.8)\}$ as an element in basic set $\{D_1, \ldots, D_n\}$ and therefore a disease by definition.

The question of whether something that is not contained in basic set $\{D_1, \ldots, D_n\}$, like haemorrhoids, could be identified as a disease is decided by the degree ε of partial similarity.

For example, $\varepsilon = 0.6$ is asked and there is a human condition like: haemorrhoids$\backslash\{(C_1, 0.9), (C_2, 0.2), (C_3, 0.55)\}$, the result is:

p-similar (haemorrhoids$\backslash X$, heart_attack$\backslash Y$) = 0.66.

Since 0.66 > 0.6 haemorrhoids can be described as disease.

According to this definition a proper choice of ε is essential: The smaller the ε that is chosen, the more diseases will exist and vice versa.

However, the value of ε is not chosen by the doctor, but by society.

Anyway, this notion of disease is a notion that can be comprised in binary logic, because there is made an explicit difference between states that are consistent with a disease and states that are not. Therefore, Sadegh-Zadeh expands this notion of disease to *a* notion of "disease to a certain degree". This can be achieved by the definition as follows:

Let's assume $\mathcal{H}$ to be a small set of human conditions. A fuzzy set $\mathcal{D}$ over $\mathcal{H}$ is considered as *a* set of diseases only if there is a subset $\{D_1, \ldots, D_n\}$ of $\mathcal{H}$ and there is a function $\mu_{\mathcal{D}}$, so that:

$\mu_{\mathcal{D}}$: $\mathcal{H}$ -> [0,1] with $\mu_{\mathcal{D}}(H_i\backslash X)$ =

- 1, if $H_i\backslash X \in \{D_1, \ldots, D_n\}$, called prototype disease
- ε, if there is a prototype disease $Hj\backslash Y$ with p-similar ($H_i\backslash X$, $H_j\backslash Y$) = ε, and there is no prototype disease $H_k\backslash Z$ with p-similar ($H_i\backslash X$, $H_k\backslash Z$) > ε and $\mathcal{D} = \{(H_i, \mu_{\mathcal{D}}(H_i)) \mid H_i \in \mathcal{H}\}$.

In this expanded definition a fuzzy set of following kind is created:

$\mathcal{D} = \{(D_1, \mu_{\mathcal{D}}(D_1)), \ldots, (D_q, \mu_{\mathcal{D}}(D_q))\}$, which consists of individual archetypes of diseases, which are all members of the set $\mathcal{D}$ to different degrees.

The membership-degree $\mu_{\mathcal{D}}(D_i)$ is of interval [0,1].

These new findings are now applied to the example of haemorrhoids:

- haemorrhoids$\backslash\{(C_1, 0.9), (C_2, 0.2), (C_3, 0.55)\}$.

These criteria are compared with a prototype disease. The already known set heart_attack is called into the equation: heart_attack$\backslash\{(C_1, 1), (C_2, 0.7, (C_3, 0.8)\}$.

Drawing a comparison shows that haemorrhoids may be considered as a disease to a degree of 0.66.

Accepting another individual with another set, haemorrhoids$\backslash\{(C_1, 0.2), (C_2, 0.1), (C_3, 0.1)\}$, and taking this individual in comparison to heart_attack would result in a membership of degree 0.16 to the set of diseases. From this, we conclude that a person may have a disease to a certain degree and that this person may have no disease to a certain degree at the same time. [1]

To demonstrate Sadegh-Zadeh's ideas of the fuzzy-disease, we implemented a computer program that we already presented in an earlier paper. [4]

By making requests of membership degrees of certain symptoms, similarities to existing diseases that are stored in a database are calculated and decisions will be made on whether an entered group of symptoms provides an indication of the existence of a disease.

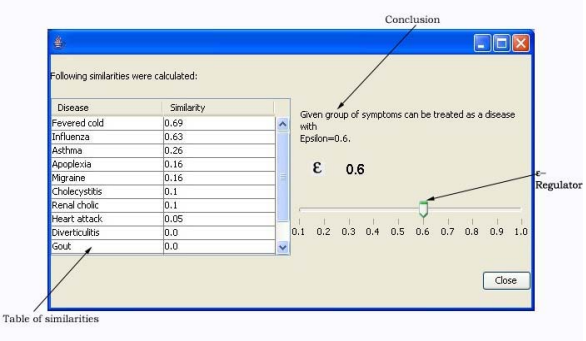

Figure 2: Form of calculation of similarities – analysis.

The preset ε-value is compared with the reference value, the highest value of similarity (listed at the top of the table of similarities). If the value of ε is higher than the reference value or is equal to this value, at least, the conclusion is drawn that the given group of symptoms can be treated as a disease to an extent of the degree ε.

An organisms attributes, that can be measured, e.g. by determine the temperature of the body, or viewed, e.g. "wheals", are called phenotypes. [5]

So far, by indicating different symptoms, we discussed the phenotype attributes of a disease. Now, our point of interest will be the cause and the roots of a disease.

According to the director of the National Institute of Environmental Health Sciences Kenneth Olden "diseases are caused by multi-factorial interactions of genes and environment". [6]

Let's conjure up the definition of diseases that Rothschuh claimed: Diseases are based on seeds. One may assume that seeds grow better if there is a fertilizer and if seeds are planted in a convenient kind of ground. So, the fertilizer can be compared to the environmental influence and the ground, as a living material, can be compared with the human body and therefore his genes. Now, it is time for a ground survey – or rather a definition of genes.

3 Genes and Fuzzy Genes

In June 27, 1994, Bill Gates quoted in *Business Week* "The gene is by far the most sophisticated program around" and in the *Stanford Encyclopedia of Philosophy* H.-J. Rheinberger declared inter alia: "There has never been a generally accepted definition of the 'gene' in genetics. There exist several, different accounts of the historical development and diversification of the gene concept as well. Today, along with the completion of the human genome sequence and the beginning of what has been called the era of postgenomics, genetics is again experiencing a time of conceptual change, voices even being raised to abandon the concept of the gene altogether." [7]

In this paper, the concept of genes will not be abandoned. Instead of that, we will keep in mind that the definition of the gene has always been changing according to newest findings in science and accept the definition of the gene, formulated by the Sequence Ontology Consortium as "a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions". According to Karen Eilbeck, a co-ordinator of the Sequence Ontology Consortium, it took 25 scientists nearly two days to reach this definition of a gene. [8]

The "locatable region of genomic sequence", respective all the genetic instructions for the development and functioning of living organisms are contained within the DNA[1]. The DNA is a double helix and made from many units of nucleotides. A nucleotide consists of a base, a sugar and one ore more phosphate groups.

In the DNA, the backbone consists of the phosphates and sugars and the purine and pyrimidine bases adenine (A), guanine (G), cytosine (C) and thymine (T) are inward-looking.

A segment of DNA may code a protein. The genetic code describes the relationship between the DNA sequence and the protein sequence. Only one of the two strands of the DNA codes the protein.

[1] As an exception, one has to mention that there is a group of viruses that have RNA-genomes.

A coded DNA sequence consists of many codons, which are read from a certain starting point.

Every codon consists of three nucleotides and accords to one amino acid. Indeed, the DNA of a gene contains all the Information that is needed for synthesis of protein, but DNA isn't the direct matrice for its creation.

In fact, the genetic information of the DNA first has to be transcribed in the base sequence of a single-stranded ribonucleic acid – the RNA, which contains of the sugar ribose and the base uracil (U) instead of 2-deoxyribose and thymine. After other transformations, a working copy of the gene is achieved, called messenger-RNA (mRNA). This mRNA describes a transportable information system for the synthesis of a specific protein [5][9][10].

So far, we've described the genesis, the location and the composition of a gene. Now, the reader might be interested in the questions of "How many genes are in the individual's organism?", respective "How to count bases?" and "How may one analyze given bases in order to recognize a disease?".

According to Ernst Peter Fisher, who is professor of the history of science at the university of Constance, the concept of genes is fuzzy. As if one only takes the sequences of the genome into account, one also has to recognize that genes are fuzzy entities. There are sequences that overlap and that recur – from those sequences the membership to a specific gene is unclear and counting is only possible if there is a definition [11]. Although there are problems in finding a precise definition and in giving an impression of the count of genes, definitions and the number of genes are still in the human's point of interest in order to verify the gene predictions.

In 1995, Victor Velculescu developed a new technique called SAGE (serial analysis of gene expression). First, RNA molecules are isolated and then their sequence is copied transcribed to DNA. Finally, one receives a piece of 20 pairs of bases from this copy of DNA; with this piece the gene that is considered may be identified [12][13]. To compare and to determine genes, one consults a database that stores information about already known sequences. The National Center for Biotechnology Information (NCBI)[2] offers a database that stores a collection of all publicly available DNA sequences that can be drawn for comparisons.

Comparing sequences is a difficult task and there are many different methods that describe possibilities and algorithms. In public databases, genes are stored as well-defined, crisp sets of bases.

As already adumbrated, there are good reasons for considering the gene in a fuzzy-theoretical way.

In the article "Fuzzy Genomes", Sadegh-Zadeh justifies the need of a fuzzy definition and shows a way of realization [14]. Furthermore, he developed a method to compare these fuzzy genomes.

The last section of this paper deals with a short summary of Sadegh-Zadeh's fuzzy genomes and another approach of sequence comparison by Angela Torres and Juan J. Nieto [15] will be taken into contrast.

4 Fuzzy genomes and comparisons on base sequences – two approaches

Both theories that will be presented are based on a fuzzy-theoretical definition of the gene and particular attention is paid on the comparison in order to identify diseases.

4.1 Approach 1: Fuzzy genomes by Sadegh-Zadeh

Having analysed a human's germplasm, one has to decide if a given section of RNA is a disease, respectively, a special form of disease, such as HIV.

Decisions on these cases are made by comparing known sequences of diseases with the section of RNA. Therefore, Sadegh-Zadeh transforms DNA and RNA to fuzzy sets.

According to the "RNA alphabet" of the bases <U, C, A, G>, U could be written as 1000, because appearance of U is true and there is no C, no A and no G. C could be written as 0100, A as 0010, G as 0001. So, an RNA sequence UACUGU can be transformed into the following bit sequence: 1000001001001000 00011000.

[2] http://www.ncbi.nlm.nih.gov/

This sequence has a length of 24 and can be represented in a 24-dimensional bit vector: (1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0).

To combine all possible appearances of a character in the alphabet, Sadegh-Zadeh builds up a fuzzy-matrix. This matrix contains all bases, or rather every character of the RNA alphabet and its membership to a given base sequence.

Considering every single base when building up a matrix, there are two points of interest:

1) Where is the position of the base and
2) To what extent and accordingly to which membership are the bases given.

For example, there is a RNA sequence UAC. Thus, the sequence consists of three bases.

The corresponding fuzzy-matrix would be:

Fuzzy_matrix (UAC) =

<(U in 1, 1), (C in 1, 0), (A in 1, 0), (G in 1, 0)

(U in 2, 0), (C in 2, 0), (A in 2, 1), (G in 2, 0)

(U in 3, 0), (C in 3, 1), (A in 3, 0), (G in 3, 0)>

An example, considering the first row: In UAC, U is at the first position. Therefore, U has membership 1, at position 1 (write U in 1, 1), there is no C at position 1, therefore C has membership 0 to the first position, the same with A and G.

One has to mention, that appearance or membership isn't obliged to be either 1 or 0. Membership may accept every value between 0 and 1. This could be the case, if a base is defective and can't be defined in one single class or there is uncertainty about the correct identification of a piece of a sequence.

Having a (m x n)-matrix, one may build up the corresponding (m x n)-vector with length n - thus, creating an n-dimensional vector.

Dealing with dimensions, Sadegh-Zadeh used the hypercube that was introduced into fuzzy set theory by Bart Kosko in [16].

For instance, there is a fuzzy set A with $\{(x_1, a_1), \ldots, (x_n, a_n)\}$. This set has an n-dimensional vector $(a_1, \ldots, a_n)$ and its members are in [0,1]. Consequently, it is a point in an n-dimensional unit hypercube $[0,1]^n$. The n appearances of x_i are assigned to coordinates in the hypercube.

According to the ground set Ω with $n \geq 1$ members, the fuzzy powerset is given as $F(2^{\Omega})$. Hence, one builds up a hypercube with 2^n corners.

If A consists of 3 members, one could display A in a 3-dimensional cube (with 2^3=8 corners).

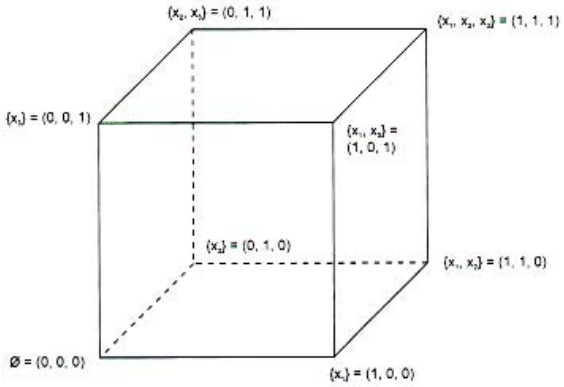

Figure 3: 3-dimensional cube [14].

For example, considering the fuzzy set A as $\{(x_1, 0.5), (x_2, 0.4), (x_3, 0.7)\}$. According to the coordinate axes x_1, x_2, x_3, A would be a point (0.5, 0.4, 0.7) in the 3-dimensional hypercube.

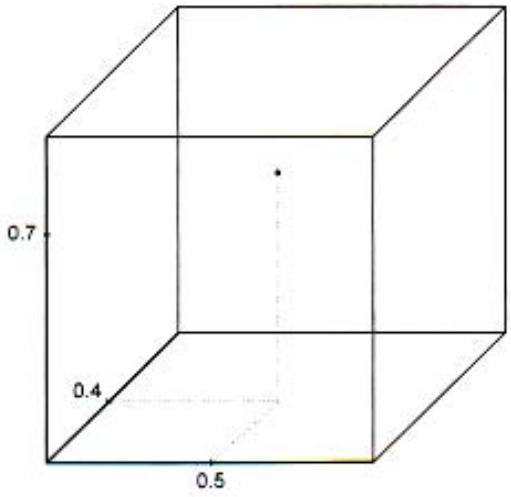

Figure 4: Point (0.5, 0.4, 0.7) in 3-dimensional cube [14].

A sequence of DNA or RNA is a point in a $4n$-dimensional hypercube. The $4n$ dimensions are reasoned by the fact that there are four bases, so m=4.

Displaying a sequence as a point in an n-dimensional hypercube depicts its order in comparison to any other sequence.

This suggests that one calculates differences or similarities between ordered fuzzy sets: Differences and, consequently, similarities of two polynucleotides A and B may be calculated after the definition of the difference, already mentioned as:

$$differ\ (A,B) = \frac{\sum_i \max(\ 0, \mu_A(x_i) - \mu_B(x_i)) + \sum_i \max(\ 0, \mu_B(x_i) - \mu_A(x_i))}{c(A \cup B)}$$

or as Hamming distances in the cube.

Analogous, similarity between sequences are calculated as the inverse of difference or as:

$$similar\,(A,B) = \frac{c(A \cap B)}{c(A \cup B)}$$

A degree that determines the vagueness of a set is referred to its fuzzy entropy, denoted as *ent*, so that the hypercube is mapped as follows:

$$ent : F(2^{\Omega}) \rightarrow [0, 1]$$

Considering a set's entropy, one is interested in determining the *nearest* and the *farthest* set.

Let's assume that there is a set A = (0.2, 0.8, 0.6). Then the nearest and farthest sets are given as: A_{near}= (0, 1, 1) and A_{far} = (1, 0, 0).

According to the hypercube, there is always a nearest and a farthest vertex to A. Fuzzy entropy of any set A is defined as the ratio of the Hamming distance from vertex A_{near} to A_{far}:

$$ent(A) = \frac{l^1(A, A_{near})}{l^1(A, A_{far})}$$

Clarity, denoted as *clar(A)*, is defined as the opposite to fuzzy entropy:

$$clar(A) = 1 - ent(A)$$

At the edges, *clar(A)* = 1 and in the centre of the hypercube *clar(A)* = 0.

From that, we can follow that real, what means existing, polynucleotides like UAC has entropy of 0 and therefore clarity of 1, whereas a fuzzy polynucleotide is near and far from a real polynucleotide to a certain degree [14].

4.2 Approach 2: The fuzzy polynucleotide space: basic properties by Torres and Nieto

As this approach is partly similar to the approach that has been presented firstly, only the common bases and the differences will be described.

Torres' and Nieto's approach bases on Sadegh-Zadeh's approach; a 12-dimensional hypercube and the RNA-Alphabet are also taken in consideration. The main difference between their theory and Sadegh-Zadeh's theory results from the fact, that Torres and Nieto do not generalize the 12 dimension to *n* dimensions. Instead of doing so, they leave at 12 dimensions and compare genes by frequency of occurrence of a certain base.

In a given sequence there are the four bases U, C, A, G and three of these bases build up a codon. A sequence is now characterized because of the frequency of every single base at any position in every codon.

For example, we consider a sequence as: UACUGA. The codons would be given as $codon_1$: UAC and $codon_2$: UGA. Considering U, we would conclude that U occurs at position 1 in $codon_1$ and also, U occurs at position 1 in $codon_2$. All in all, U occurs 2 times in the whole sequence at position 1. Thus, the fraction of U in the first base is calculated as 2/2 = 1 = 100%.

By applying this method to every base, 3*4 = 12 fractions will be calculated, as there are 3 positions in a codon and 4 possible bases.

The following table shows a table of fractions of a sequence S_1 that would be given for example as: S_1 = CAUUGU

Table 1: Table of fractions of sequence CAU UGU.

Position	**Count of nucleotides**					**Fractions**			
	U	C	A	G	*Sum*	U	C	A	G
1.	1	1	0	0	*2*	0.5	0.5	0	0
2.	0	0	1	1	*2*	0	0	0.5	0.5
3.	2	0	0	0	*2*	1	0	0	0

After calculating fractions of a base, a vector of fractions with length = 12 remains and stands for the whole sequence. In the example above, the sequence S_1 with CAU UGU would result in a vector V_1 = {0.5, 0.5, 0, 0, 0, 0, 0.5, 0.5, 1, 0, 0, 0}.

In order to compare a sequence with another sequence, the sequences' vectors of fractions are considered and the similarity between these sequences is calculated as:

$$sim(A, B) = c(A \vee B) / C$$

with

$$C = \left(\frac{a_1 + b_1}{2}, \ldots, \frac{a_n + b_n}{2} \right)$$

The difference between sequences is given as:

$$dif(A, B) = 1 - sim(A, B)$$

They conclude that:

- *sim(A,B) ≠ similar(A,B)* and
- *dif(A,B) ≠ differ(A,B)*.

Every sequence of bases can be compared with every other sequence, by comparing the 12 fractions of the sequences, whether they are of the same length or not [15].

Acknowledgement

This work is partially supported by Grant # AB00158OFF from the Vienna Federal Data Center.

References

[1] Kazem Sadegh-Zadeh: Fuzzy Health, Illness, and Disease. Journal of Medicine and Philosophy, Vol. 25, No. 5 (2000), pp. 605-638: 612

[2] Kazem Sadegh-Zadeh: Fundamentals of clinical methodology: 3. Nosology. Artifical Intelligence in Medicine, 17 (1999) 81-108

[3] Karl Eduard Rothschuh: Der Krankheitsbegriff (Was ist Krankheit?), Hippokrates 43 (1972) 3-17

[4] Julia Limberg and Rudolf Seising: Fuzzy Health, Illness, and Disease Sadegh-Zadeh's framework and a program to identify diseases. Proceedings of the FUZZ-IEEE 2007, IEEE International Conference on Fuzzy Systems :: Intelligence is Fuzzy::, Imperial College, London, UK, 23-26 July, 2007 (in print).

[5] Benjamin Lewin: *Molekularbiologie der Gene*. Spektrum Akademischer Verlag. 1998.

[6] OLPA,107th Congress, Hearing I: Testimony on Environmental Health Problems. Available at: http://olpa.od.nih.gov/hearings/107/session2/testimonies/environmental.asp

[7] Hans-Jörg Rheinberger: Gene. Stanford Encyclopedia of Philosophy. Available at: http://plato.stanford.edu/entries/gene/

[8] Neurophilosophy. The Concept Of A Gene Has Envolved. Available at: http://neurophilosophy.wordpress.com/2006/05/30/the-concept-of-a-gene-has-evolved/

[9] Gerhard Thews, Ernst Mutschler, Peter Vaupel: *Anatomie, Physiologie, Pathophysiologie des Menschen*. Stuttgart: Wissenschaftliche Verlagsgesellschaft, Auflage: 5. 1999.

[10] DNA: http://en.wikipedia.org/wiki/DNA

[11] Ernst Peter Fisher: Gene sind wie Sand am Meer. Available at: http://www.netdoktor.de/feature/gene_fischer.htm

[12] SAGE: http://en.wikipedia.org/wiki/Serial_Analysis_of_Gene_Expression

[13] Ernst Peter Fischer: *Das Genom*. Frankfurt: Fischer (Tb.). Auflage: 2., 2002.

[14] Kazem Sadegh-Zadeh: Fuzzy Genomes. Artificial Intelligence in Medicine, 18 (2000) pp. 1-28.

[15] Angela Torres and Juan J. Nieto: The fuzzy polynucleotide space: basic properties. In: Bioinformatics, Vol.19, No.5 (2003), pp.587-592

[16] Bart Kosko: Fuzzy Entropy and Conditioning, Information Sciences, 40 (1986), pp. 165-174.