*Sitanath Biswas International Journal of Engineering Research and Application www.ijera.com*
*ISSN: 2248-9622 Vol. 5, Issue 12 (Series -4) Dec. 2015, pp 107-110*

RESEARCH ARTICLE OPEN ACCESS

# Fuzzy Logic in Medicine and Bioinformatics

Amitav Saran1, Subhashree Sukla 2, Prabakaran P 3
*[1,2]Associate Professor, Department of Computer Science Engineering, Gandhi Institute For Technology (GIFT), Bhubaneswar*
*[3] Assistant Professor, Department of Computer Science Engineering, Gandhi Engineering College, Bhubaneswar*

**ABSTRACT;** The purpose of this paper is to present a general view of the current applications of fuzzy logic in medicine and bioinformatics. We particularly review the medical literature using fuzzy logic.We then recall the geometrical interpretation of fuzzy sets as points in a fuzzy hypercube and present two concrete illustrations in medicine (drug addictions) and in bioinformatics (comparison of genomes).

## INTRODUCTION

The diagnosis of disease involves several levels of uncertainty and imprecision, and it is inherent to medicine. A single disease may manifest itself quite differently, depending on the patient, and with different intensities. A single symptom may correspond to different diseases. On the other hand, several diseases present in a patient may interact and interfere with the usual description of any of the diseases. The best and most precise description of disease entities uses linguistic terms that are also imprecise and vague.Moreover, the classical concepts of health and disease are mutually exclusive and opposite. However, some recent approaches consider both concepts as complementary processes in the same continuum [1–6]. According to the definition issued by the World Health Organization (WHO), health is a state of complete physical, mental, and social well-being, and notmerely the absence of disease or infirmity. The loss of health can be seen in its three forms: disease, illness, and sickness.

To deal with imprecision and uncertainty, we have at our disposal fuzzy logic. Fuzzy logic introduces partial truth values, between true and false. According to Aristotelian logic, for a given opposition or state we only have two logical values: true-false, black-white, 1-0. In real life, things are not either black or white, but most of the times are grey. Thus, in many practical situations, it is convenient to consider intermediate logical values. Let us show this with a very simple medical example. Consider the statement "you are healthy." Is it true if you have only a broken nail? Is it false if you have a terminal cancer? Everybody is healthy to some degree h and ill to some degree i. If you are totally healthy, then of course h = 1, i = 0. Usually, everybody has some minor health problems and h <1, but h + i = 1. (1)

In the other extreme situation, h = 0, and i = 1 so that you are not healthy at all (you are dead). In the case you have only a broken nail, we may write h = 0.999, i = 0.001; if you have a painful gastric ulcer, i = 0.6, h = 0.4, but in the case you have a terminal cancer, probably i = 0.95, h = 0.05. As we will see, this is a particular case of Kosko's hypercube: the one-dimensional case [4]. Uncertainty is now considered essential to science and fuzzy logic is a way to model and deal with it using natural language. We can say that fuzzy logic is a qualitative computational approach.

Since uncertainty is inherent in fields such as medicine and massive data in bioinformatics, and fuzzy logic takes into account such uncertainty, fuzzy set theory can be considered as a suitable formalism to deal with the imprecision intrinsic to many biomedical and bioinformatics problems.
Fuzzy logic is a method to render precise what is imprecise in the world of medicine. Several examples and illustrations are mentioned below.

## FUZZY LOGIC INMEDICINE

The complexity ofmedical practice makes traditional quantitative approaches of analysis inappropriate. In medicine, the lack of information, and its mprecision, and, many times, contradictory nature are common facts. The sources of uncertainty can be classified as follows [7].

(1) Information about the patient.
(2) Medical history of the patient, which is usually supplied by the patient and/or his/her family. This is usually highly subjective and imprecise.
(3) Physical examination. The physician usually obtains objective data, but in some cases the boundary between normal and pathological status is not sharp.
(4) Results of laboratory and other diagnostic tests, but they are also subject to some mistakes, and even to improper behavior of the patient prior to the examination.

(5) The patient may include simulated, exaggerated, understated symptoms, or may even fail to mention some of them.
(6) We stress the paradox of the growing number of mental disorders versus the absence of a natural classification[8].\
The classification in critical (ie, borderline) cases is difficult, particularly when a categorical system of diagnosis is considered. Fuzzy logic plays an important role in medicine [7, 9–14]. Some examples showing that fuzzy logic crosses many disease groups are the following.
(1) To predict the response to treatment with citalopram in alcohol dependence [15].
(2) To analyze diabetic neuropathy [16] and to detect early diabetic retinopathy [17].
(3) To determine appropriate lithium dosage [18, 19].
(4) To calculate volumes of brain tissue from magnetic resonanceimaging (MRI) [20], and to analyze functionalMRI data [21].
(5) To characterize stroke subtypes and coexisting causes of ischemic stroke [1, 3, 22, 23].
(6) To improve decision-making in radiation therapy [24].
(7) To control hypertension during anesthesia [25].
(8) To determine flexor-tendon repair techniques [26].
(9) To detect breast cancer [27, 28], lung cancer [28], or prostate cancer [29].
(10) To assist the diagnosis of central nervous systems tumors (astrocytic tumors) [30].
(11) To discriminate benign skin lesions from malignant melanomas [31].
(12) To visualize nerve fibers in the human brain[32].
(13) To represent quantitative estimates of drug use[33].
(14) To study the auditory P50 component in schizophrenia[34].
(15) Many other areas of application, to mention a few, are
(a) to study fuzzy epidemics [35],
(b) to make decisions in nursing [36],
(c)to overcome electro acupuncture ac commodation[37].

We used the database MEDLINE to identify the medical publications using fuzzy logic. We used as keywords fuzzy logic and grade of membership. The total number of articles per year appears inTable 1. The data is from 1991 to 2002 and includes also the number of those publications in 1990 and before. It results in a total of 804 articles and agrees essentially with the numbers indicated in [7, 13].We plan to creen databases in the engineering literature that covers medicine related articles since it is difficult to publish medical results using a fuzzy logic approach. In the future we will compare the figures obtained.

Figure 1 indicates an exponential growth in the number of articles in medicine making use of fuzzy technology. The preliminary data we have for 2003 and 2004 [38] supports this tendency.

**FUZZY LOGIC IN BIOINFORMATICS**

Bioinformatics derives knowledge from computer analysis of biological data. This data can consist of the information stored in the genetic code, and also experimental results (and hence imprecision) from various sources, patient statistics, and scientific literature. Bioinformatics combines computer science, biology, physical and chemical principles, and tools for analysis and modeling of large sets of biological data, the managing of chronic diseases, the study of molecular computing, cloning, and the development of training tools of bio-computing systems [39]. Bioinformatics is a very active and attractive research field with a high impact in new technological development [40].

Molecular biologists are currently engaged in some of the most impressive data collection projects. Recent genome sequencing projects are generating an enormous amount of data related to the function and the structure of biological molecules and sequences. Other complementary high-throughput technologies, such as DNA microarrays, are rapidly generating large amounts of data that are too overwhelming for conventional approaches to biological data analysis. We have at our disposal a large number of genomes, protein structures, genes with their corresponding
Expressions monitored in experiments, and single-nucleotide polymorphisms (SNPs) [41].

For example, the EMBL Nucleotide Sequence Database (http://www.ebi.ac.uk/embl) has increased in 12 months from 18.3 million entries comprising 23 Gb (Release 71, September 2002) to 27.2 million entries comprising over 33 Gb (Release 76, September 2003) as indicated in [42].

Handling this massive amount of data, in many cases imprecise and fuzzy, requires powerful integrated bioinformatics systems and new technologies.

Fuzzy logic and fuzzy technology are now frequently used in bioinformatics. The following are some examples.
(1) To increase the flexibility of protein motifs [43].
(2) To study differences between polynucleotides [44].
(3) To analyze experimental expression data [45] using fuzzy adaptive resonance theory.

(4) To align sequences based on a fuzzy recast of a dynamic programming algorithm [46].
(5) DNA sequencing using genetic fuzzy systems [47].
(6) To cluster genes from microarray data [48].
(7) To predict proteins subcellular locations from their dipeptide composition [49] using fuzzy k-nearest neighbors algorithm.
(8) To simulate complex traits influenced by genes with fuzzy-valued effects in pedigreed populations [50].
(9) To attribute cluster membership values to genes [51] applying a fuzzy partitioning method, fuzzy C-means.
(10) To map specific sequence patterns to putative functional classes since evolutionary comparison leads to efficient functional characterization of hypothetical proteins [52]. The authors used a fuzzy alignment model.
(11) To analyze gene expression data [53].
(12) To unravel functional and ancestral relationships between proteins via fuzzy alignment methods [54], or using a generalized radial basis function neural network architecture that generates fuzzy classification rules [55].
(13) To analyze the relationships between genes and decipher a genetic network [56].
(14) To process complementary deoxyribonucleic acid (cDNA) microarray images [57]. The procedure should be automated due to the large number of spots and it is achieved using a fuzzy vector filtering framework.
(15) To classify amino acid sequences into different super families [58].

## THE FUZZY HYPERCUBE

In 1992, Kosko [4] introduced a geometrical interpretation of fuzzy sets as points in a hypercube. In 1998, Helgason and Jobe [1] used the unit hypercube to represent concomitant mechanisms in stroke. Indeed, for a given set X = _x1, . . . , xn_, (2)

a fuzzy subset is just a mapping $\mu : X \longrightarrow I = [0, 1]$, (3) and the value $\mu(x)$ expresses the grade of membership of the element $x \in X$ to the fuzzy subset $\mu$.

For example, let X be the set of persons of some population and let the fuzzy set $\mu$ be defined as healthy subjects. If John is a member of the population (the set X), then, $\mu$ (John) gives the grade of healthiness of John, or the grade of membership of John to the set of healthy subjects. If $\lambda$ is the fuzzy set that describes the grade of depression, then $\lambda$ (Mary) is the degree of depression of Mary.

Thus, the set of all fuzzy subsets (of X) is precisely the unit hypercube $I^n = [0, 1]^n$, as any fuzzy subset $\mu$ determines a point $P \in I^n$ given by $P = (\mu(x1), \ldots, \mu(xn))$. Reciprocally, any point $A = (a1, \ldots, an) \in I^n$ generates a fuzzy subset $\mu$ defined by $\mu(xi) = ai$, $i = 1, \ldots, n$. Nonfuzzy or crisp subsets of X are given by mappings $\mu : X \rightarrow \{0, 1\}$, and are located at the $2^n$ corners of the n-dimensional unit hypercube $I^n$. For graphic representations of the two-dimensional and threedimensional hypercube, we refer to [59].
Given, p = _p1, p2, . . . , pn_, q = _q1, q2, . . . , qn_ $\in I^n$, (4) not both equal to the empty set$\emptyset$ = (0, 0, . . . , 0), we define the difference between p and q as
Of course $d(\emptyset, \emptyset) = 0$.We know that d is indeed a metric [60]. Hypercubical calculus has been described in [61], while some biomedical applications of the fuzzy unit hypercube are given in [1, 6, 59]. Recently, the fuzzy hypercube has been utilized to study differences between polynucleotides [59] and to compare genomes [44, 62].

## AN APPLICATION TO DRUG ADDICTIONS

We now present an example of the use of the fuzzy hypercube in a medical case of consumption of drugs.

Consider the following fuzzy variables: smoking and alcohol drinking. If you do not smoke, then your degree of being a smoker is evidently 0. If you smoke, for example, six cigarettes per day, we say that your degree of being a smoker is 0.8. If the consumption is ten or more, the degree is 1. See [63, Figure 3.8] for a geometrical representation of the fuzzy concept of being a smoker. With respect to the other fuzzy variable, if you drink no alcohol, the degree of this variable is 0. If you drink more than 75 cc of alcohol per day, the degree of alcoholism is 1.
For 25 cc/d, the degree could be 0.4 and for 50 cc/d, 0.8.

Thus, the fuzzy set $\mu = (0, 0)$ corresponds to a nonsmoker and teetotaler. Some further examples are the following: the set $\mu = (1, 0)$ represents a heavy smoker, but a teetotaler, and the set $\mu = (0.8, 1)$ is a person who smokes about six cigarettes a day and is a risk consumer of alcohol. Suppose you correspond to the fuzzy set $\lambda = (1, 1)$, have recently had some health problems, and your physician has advised you to reduce your consumption of cigarettes and alcohol by half. The ideal situation for your health is, of course, the point $\mu = (0, 0)$, but it is possibly difficult to achieve. Cigarette smoking and alcohol drinking during adolescence have been shown to be associated with a greater possibility of concurrent and future substance-related disor-ders (Lewinsohn et al [64]; Nelson andWittchen [65]).

In order to report patterns of drug use and to describe factors associated with substance use in adolescents, a cross-sectional survey was carried out in a representative population sample of 2550 adolescents, aged 12 to 17 years, from Galicia (an autonomous region located in the Northwest of Spain). The original survey covered the use of alcohol, tobacco, illicit drugs, and other psychoactive substances. For tobacco smoking and alcohol drinking, each subject of the population sample was assigned a fuzzy degree of addiction (or risk use) and mapped into the two-dimensional hypercube I2 by
an expert.

Several subjects occupy the same point in the two dimensional hypercube. For example Figure 2 represents the number of subjects in the cross-sectional survey according to the two fuzzy degrees of addiction. The reader can see that there are 1278 subjects corresponding to the point (0, 0), that is, nonsmoker and teetotaler. Also 7 adolescents are at the point (0.8, 0.2). There are 121 subjects on the line of probability $x1 + x2 = 1$.

Indeed (see Figure 2), $23 + 1 + 1 + 2 + 2 + 7 + 1 + 84 = 121$.

Most subjects were inside the hypercube but outside the line of probability. This means that the vast majority of subjects (2429/2550 ≈95.25%) are outside the line of probability.

This is in agreement with the fundamental limitation of probability theory with respect to clinical science in general [1] and agrees with its results (29/30 ≈96.66%). We refer to [59] for details on the general theory of fuzzy midpoints and their applications. It has been used recently to average biopolymers [66].

## AN APPLICATION TO THE COMPARISON OF GENOMES

Whole genome sequence comparison is important in bioinformatics [44, 67]. The complete genome sequence of Mycobacterium tuberculosis H37Rv is available at http://www.ncbi.nlm.nih.gov with accession number NC 000962. The genome comprises 4 411 529 base pairs, contains around 4000 genes, and has a very high guanine + cytosine content [68]. Computing [44] the number of the nucleotides at the three base sites of a codon in the coding sequences of M tuberculosis (Table 2), and then calculating the corresponding fractions, we have the fuzzy set of frequencies of the genome sequence of M tuberculosis (Table 3).

This set can be considered as a point in the hypercube I12. Indeed, the point (0.1632, 0.3089, 0.1724, 0.3556, 0.2036, 0.3145, 0.1763, 0.3056, 0.1645, 0.3461, 0.1593, 0.3302) ∈ I12. (6)

Aquifex aeolicus was one of the earliest diverging, and is one of the most thermophilic, bacteria known [69]. It can grow on hydrogen, oxygen, carbon dioxide, and mineral salts. The complex metabolic machinery needed for A aeolicus to function as a chemolithoautotroph (an organism which uses an inorganic carbon source for biosynthesis and an inorganic chemical energy source) is encoded within a genome that is only one-third the size of the E coli genome.

The corresponding data for A aeolicus was obtained from http://www.ncbi.nlm.nih.gov with accession number NC 000918, and is presented in Tables 4 and 5, respectively. The complete genome sequence has 1 551 335 base pairs. The fuzzy set of frequencies of the genome of A aeolicus is (0.1706, 0.1605, 0.3241, 0.3446, 0.3282, 0.1735, 0.3478, 0.1504, 0.2139, 0.2455, 0.3052, 0.2352) ∈ I12. (7)

Using the distance given in (5), it is possible to compute the distance between these two fuzzy sets representing the frequencies of the nucleotides of A aeolicus and M tuberculosis: d(A aeolicus, M tuberculosis) = 2.2125 6.106 ≈0.3623. (8) In [44] we calculate the difference between Mtuberculosis and E coli K-12 obtaining d(M tuberculosis, E coli) = 0.8506 3.4253 ≈0.2483. (9) d(A aeolicus, E coli) = 0.8514 5.0161 ≈0.1697. (10)

## REFERENCES

[1]. Jobe TH, Helgason CM. The fuzzy cube and causal efficacy: representation of concomitant mechanisms in stroke. Neural Networks. 1998;11(3):549–555.
[2]. Helgason CM, Jobe TH. Perception-based reasoning and fuzzy cardinality provide direct measures of causality sensitive to initial conditions in the individual patient (Invited paper). International Journal of Computational Cognition. 2003;1(2):70–104.
[3]. Helgason CM, Malik DS, Cheng S-C, Jobe TH, Mordeson JN. Statistical versus fuzzy measures of variable interaction in patients with stroke. Neuro epidemiology. 2001; 20(2):77–84.
[4]. Kosko B. Neural Networks and Fuzzy Systems. Englewood Cliffs, NJ: Prentice-Hall; 1992.
[5]. Kosko B. Fuzzy Thinking: The New Science of Fuzzy Logic. New York, NY: Hyperion Press; 1993.
[6]. Sadegh-Zadeh K. Fundamentals of clinical methodology: 3.Nosology. Artificial Intelligence in Medicine. 1999;17(1):87–108.