

Store Data from Experiments with Microorganisms Used in Food Industry

Lena Kostadinova-Georgieva^a, Atanaska Bosakova-Ardenska^{*a}, Iliyan Dobrev^b

Received 08th July 2013, Accepted 16th July 2013

Abstract: The aim of this paper is to present results from collaboration of computer engineers and experimenters in microbiology working with molecular-genetic methods. The experimenters in microbiological laboratory at the University of Food Technologies use ARDRA (Amplified Ribosomal DNA Restriction Analysis) analyses and DNA sequencing processed with BLAST (Basic Local Alignment Search Tool) algorithm to identify some microorganisms. Their results have been accumulated in designed database. This will improve the effectiveness and productivity of the molecular-genetic analyses in department of microbiology.

Keywords: DNA analysis, DNA sequence, ARDRA, Database.

1. Introduction

Microbiological analysis of food can be conducted through different methods and the most advanced ones are the molecular-genetic methods. One of them is sequencing a DNA region or gene of a given microorganism, which is subsequently compared to the DNA of reference microbial culture [1]. The sequence is a series of the letters A, C, G and T, which represents the four nucleotide bases in the DNA chain – adenine, cytosine, guanine and thymine. The letters are usually presented next to each other without intervals, e.g. the sequence AAAGTCTGAC. Each series of nucleotides larger than or equal to 4 can be called a sequence. It can be forward or reverse according to its origin form the DNA molecule. The results from the analysis of a sequence are stored in several files of different type. Usually the file, containing the sequence itself is a text file and his name contains information whether the sequence is forward or reverses (R or 27F).

Another commonly applied method for molecular-genetic analysis is ARDRA (Amplified Ribosomal DNA Restriction Analysis). It employs enzyme multiplication of the gene coding the 16S region of the ribosomal DNA and its enzyme restriction. As a result a strain specific ARDRA profile for each microorganism is derived. It includes information about the number of nucleotide bases, contained in the obtained through restriction DNA sequences.

2. Microbiological Analysis

2.1. DNA sequencing and BLAST

BLAST (Basic Local Alignment Search Tool) [2, 3] is an algorithm for finding regions of similarity between biological sequences.

^a University of Food Technologies, department of Computer Systems and Technologies, Plovdiv, Bulgaria

^b University of Food Technologies, department of Microbiology, Plovdiv, Bulgaria

* Corresponding Author: Email: abosakova@yahoo.com;
Phone.: +35932603860

Using BLAST in GenBank [3] the explorers receives four files. For example if we have a sequence with 1102 bases then we receive one text file with bases, one pdf file with signals of different bases in DNA sequence and two files which can be used for visualization of same information in web system. On figure 1 is shown a part of pdf file of one experiment. Using BLAST shows that sequence is recognized as part of genome of microorganism *Lactobacillus casei*.



Figure 2. A transmission line of a power system network.

2.2. ARDRA analysis

The ARDRA analysis is very perspective method for identifying microorganisms. Many researchers works in this field [4, 5, 6]. But now it is not known database which allows thru the internet use ARDRA results to helps some experiments. Both analysis – DNA sequencing processed with BLAST and ARDRA is very useful like express methods for identification of microorganisms in food industry and other fields of human activities. Figure 2 shows results from ARDRA analysis [4]. For example *Lactobacillus*

helveticus (number 16) has a profile with three parts with sizes 325, 194 and 119 bases and Lactobacillus plantarum (number 17) has a profile with two parts sized 300 and 259 bases.

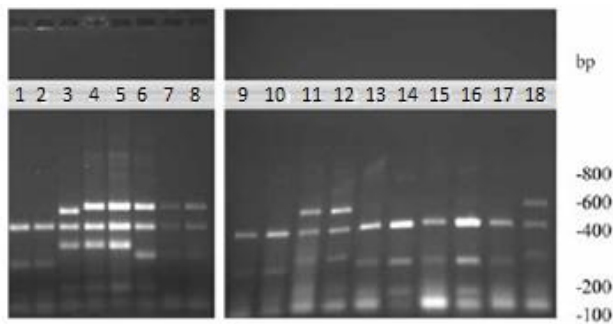


Figure 2. ARDRA profile of Lactobacillus strains obtained by digestion of amplified 16srDNA with HaeIII

3. Aim

The accumulated results from sequence analyses in the department of Microbiology at the University of Food Technologies were obtained through the application of the algorithm for sequence comparison BLAST. The aim of the designed database is to facilitate the organization and interpretation of this data, as well as the results from ARDRA analyses. This will improve the effectiveness and productivity of the molecular-genetic analyses for microbiological applications.

4. Design of Database

4.1. Design of ER model

When designing a database so-called ER model (Entity Relationships Model) is used. After analyzing the work in the laboratory the following objects have been identified:

- 1) Experiment is carried out by a particular experimenter at a specified date and is characterized by a short description and name (label);
- 2) ARDRA analysis: analysis that can be performed as part of the work on an experiment and give information about profile obtained through the use of a specific restriction enzyme;
- 3) Operator: experimenter who has a name and occupies a specific position;

- 4) Sequence (DNA sequence): naming sequence of nucleotide bases (part of the genome of the organism) received on a certain date and stored in a file;
- 5) Recognition: probability of identifying the received sequence as part of genome of a known microorganism.

Figure 3 shows the designed database ER model.

Objects shown in Figure 3 are related as follows:

- The experiment is conducted by an operator and may include results of ARDRA analysis and/or DNA sequences of organisms studied;
- The operator can perform experiments and occupies a specific position;
- ARDRA analysis is part of a concrete experiment, and the result of using specific restriction enzymes;
- The sequence was obtained in a concrete experiment, a type of segment of DNA, which is extracted and can be recognized as part of the genome of some organisms;
- Recognition is performed for a particular sequence by using an algorithm for detection in an existing genetic base and to show what is the probability the test sequence to be part of the genome of a known microorganism.

4.2. Design of Relational model

ER model is transformed into a relational model [7], consisting of a set of tables (relations) and relations between them. Figure 4 shows the structure of the relational database model. As DBMS (Database management system) is used MS Access 2007. The database is normalized to the first and second normal form [7]. The third normal form is not applied. In table Sekvencii we have two columns for which the values are connected. These columns are: Sequence and NumBase. The length of string in column Sequence is equal to number in column NumBase. But this is necessary for controlling correct input of sequences. We can check whether length of string in Sequence is equal to number in NumBase. If this is not true then we will search correct information in source files.

5. User Interface

Along with designing the database a graphical user interface (GUI) with tools provided by MS Access 2007 is also designed. Figure 5 shows the list of the developed forms.

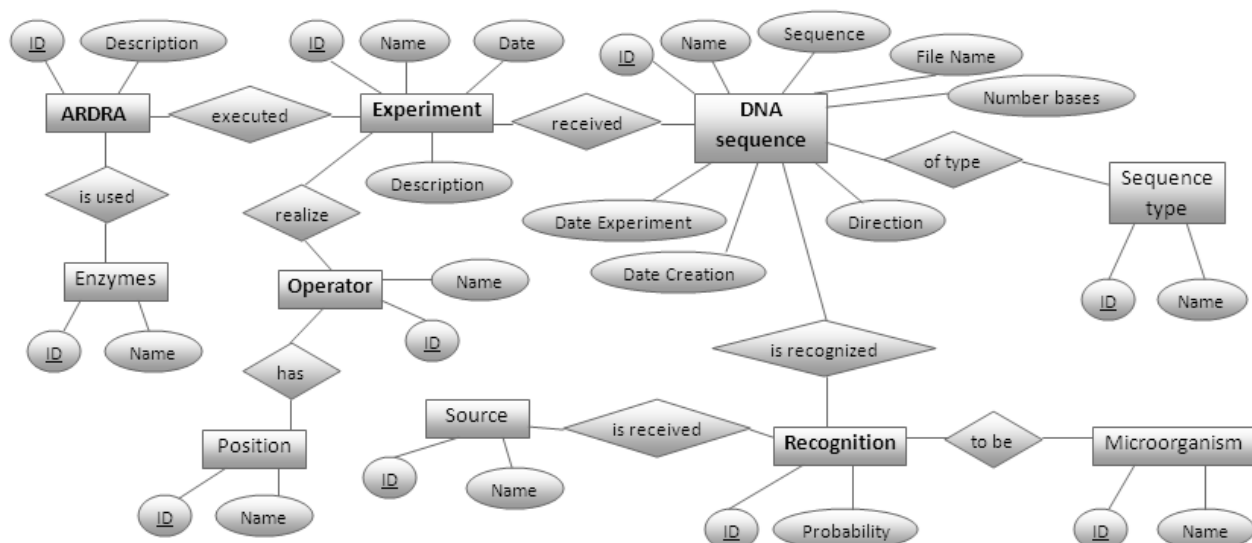


Figure 3. Entity Relationships model

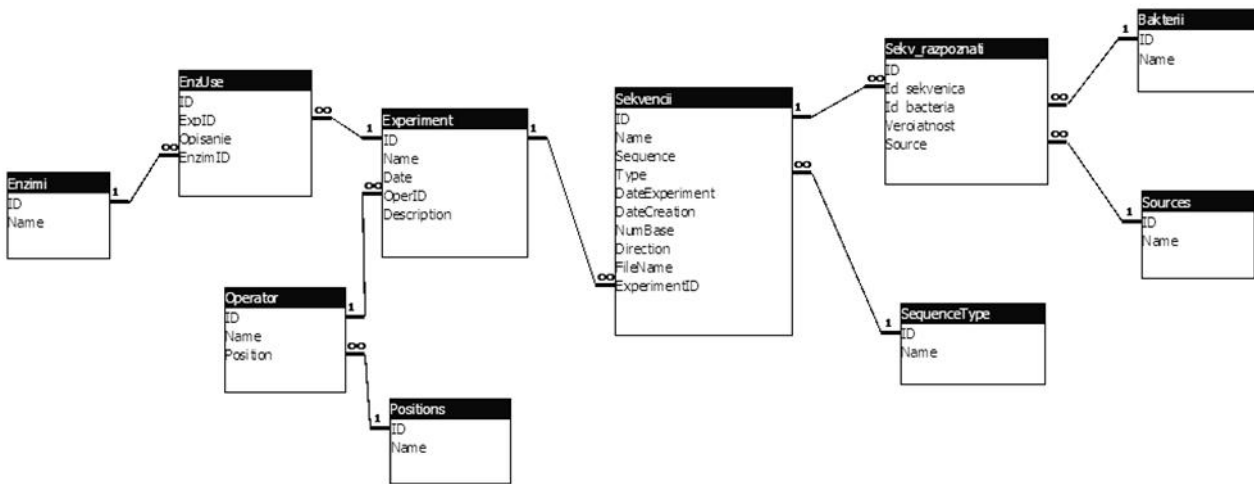


Figure 4. Relational model

When one opens the ACCDE a database file is loaded in a form Home. It contains a list of operators and after being selected an operator may consider information about conducted experiments (Fig. 6). Each experiment has a name, description and start date.

- Information about selected by user sequence;
- ARDRA analysis for a selected experiment;
- Recognitions that are made for a specific sequence.



Figure 5. GUI - List of forms

Figure 7. Form "Add DNA Sequence"

Figure 6. Form "Experiments"

For each experiment can be viewed and edited as DNA sequences also ARDRA analyzes. Figure 7 shows a form for adding sequences.

The sequence can be recognized as part of the genome of a known microorganism and this information is being edited through the forms shown in Figure 8.

Reports have been created to be used for printing the following views:

- Experiments made by the selected operator;
- Sequences associated with specific experiment;

Figure 8. Forms "Identified microorganisms" and "Recognize sequence"

The advantage of the created database is that it allows aggregation of results of analysis of sequences obtained using different algorithms [8].

In addition, for each experiment can be performed and ARDRA analysis that complement the information obtained by using algorithms for comparing in the existing genetic databases. Figure 9 shows report for ARDRA results of one experiment.

ARDRA

Experiment: Experiment Rosi

Description	Enzymes
using ARDRA - 400/500 kB	enzim
400/300 kB	enzim
using ARDRA analysis receive: 300/700 kB	enzim 2
345/120 s korekcii 20/35	enzim 3
345/250 s korekcii 20/35	enzim 3

Figure 9. Report of ARDRA results

6. Conclusion And Future Work

A database for the microbiological laboratory has been designed and implemented. The structure of the database enables to store information for identification of microorganisms obtained from various sources, and also the results from ARDRA analysis. This will significantly increase the efficiency and productivity of work with molecular genetic techniques used in microbiological practice. The work of the team continues in the following areas:

- improving the user interface of the database;
- accumulation of experimental data;

- development of an effective algorithm for detection of DNA sequences;
- publishing database in Internet.

References

- [1] I. Murgov, Z. Denkova, Genetics and selection of microorganisms, UFT Academic Publishing House, Plovdiv, 2007 (in Bulgarian)
- [2] <http://en.wikipedia.org/wiki/BLAST>
- [3] <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- [4] T. Stefanova, Urshev Z., Dimitrov Z., Fatchikova N., Minkova S. Species affiliation of dairy lactobacilli with angiotensin converting enzyme inhibitory activity. *Biotechnol. & Biotechnol. Eq.* 23, pp. 1250-1254, 2009.
- [5] I. Dobrev, Application of lactobacilli and bifidobacteria in the production of healthy foods from goat milk, PhD thesis, University of Food Technologies, Plovdiv, 2011
- [6] Petrov K., Urshev Z., Petrova P. L(+)-Lactic acid production from starch by a novel amylolytic *Lactococcus lactis* subsp. *lactis* B84. *Food Microbiology*, vol. 25, pp. 550-557, 2008.
- [7] Michael J. Hernandez, Database Design for Mere Mortals: A Hands-On Guide to Relational Database Design, Addison-Wesley Professional, 2003
- [8] Li L, Khuri S. A comparison of DNA fragment assembly algorithms. In: International conference on mathematics and engineering techniques in medicine and biological sciences; pp. 329-35, 2004.

A Note on Entropy Subsethood Relationship

Mamoni Dhar^{*a}

Received 14th May 2013, Accepted 06th August 2013

Abstract: We comment on subsethood measures defined by Kosko and Young and give some new aspects of these measures. Finally we would like to discard the entropy subsethood relationship established by the authors. We present some properties of subsethood measure from set theoretic approach and also from axiomatic approach with the expectation that these would help in removing the shortcomings that currently exist in these definitions.

Keywords: Membership value, Membership function, Entropy.

1. Introduction

Since fuzzy set was introduced by Zadeh [1], in 1965, inclusion measure and fuzzy entropy have become two important topics in fuzzy set theory and have successfully been applied in many fields such as image processing, fuzzy neural networks, fuzzy reasoning and fuzzy control.

Inclusion measure of fuzzy sets indicates the degree to which a fuzzy set A is contained in another fuzzy set B. Zadeh [1] first gave the definition of fuzzy set inclusion in a crisp relation, in another word, a fuzzy set A is either included or not included in fuzzy set B. This is defined as follows

A fuzzy set A in the Universe U is a subset of another fuzzy set B if for every element x in U, its membership degree in A is less than or equal to the membership degree in B. This can formally be stated as

$$A \subseteq B \Rightarrow \mu_A(x) \leq \mu_B(x) \quad (1)$$

The graphical representation of this definition is, he defines the fuzzy set A as a subset of another fuzzy set B if graph of A never goes above the graph of B.

Later on it was realized that defining fuzzy subsethood in this way is though highly appreciable and useful but it still against the spirit of fuzzy set theory in the sense that it presents a crisp decision about being a subset or not. For many researchers working in the area of fuzzy subsethood considered this definition to be too rigid and so remained very interested in assigning a degree of inclusion of one fuzzy set into another. As a result more variant and expressions of fuzzy subsethood for two fuzzy sets were studied (see for example [9-11]).

It is known that Kosko [2] and Young [3] had considered the connection between fuzzy subsethood and entropy of fuzzy sets. It is for this reason in this article, we shall discuss about the two most commonly used subsethood measures of which one is set theoretic approach given by Kosko and the other is the axiomatization principle of Young.

Kosko criticizes the definition of fuzzy set containment given by

Zadeh, pointing out that if this inequality hold for all but for just a few x, we can still consider A to be subset of B to some degree. He proposed a definition of subsethood of two fuzzy sets from the point of view of set theoretic approach and this is as follows:

$$S(A, B) = \frac{M(A \cap B)}{M(A)} \quad (2)$$

where S(A,B) stands for the degree of subsethood of A in B and M(A) and M(A ∩ B) are the cardinalities of the fuzzy sets A and A ∩ B respectively which are defined in the following way

$$M(A) = \sum_i \mu_A(x_i), \forall x_i \in X \quad (3)$$

and

$$M(A \cap B) = \sum_i \mu_{A \cap B}(x_i), \forall x_i \in X \quad (4)$$

After that Sinha and Dougherty [4], introduced axiomatic definition of inclusion measure of fuzzy sets.

On the basis of Kosko's subsethood measure, fuzzy entropy and Willmott's work, Young defined the concept of subsethood measure and weak subsethood measure, which can be summarised as follows:

A real function $c: F(X) \times F(X) \rightarrow [0,1]$ is called a subsethood measure, if c has the following properties

$$(C1) C(A,B)=1 \text{ if and only if } A \subseteq B, \text{ i.e. } \mu_A(x) \leq \mu_B(x)$$

$$(C2) \text{ If } \left[\frac{1}{2} \right] \subseteq A \text{ then } c(A, A^c) = 0, \text{ if and only if } A=X.$$

$$(C3) \text{ If } A \subseteq B \subseteq C \text{ then } c(C, A) \leq c(B, A) \text{ and if } A \subseteq B, \text{ then } c(C, A) \leq c(C, B)$$

He further mentioned that if c satisfies (C2) and (C3), then c is called a weak subsethood measure,

Moreover, the Kosko's subsethood measure complies with Young's axiomatic characterization of a fuzzy subsethood measure but does not fit into the framework of Sinha and Dougherty [4].

Entropy of fuzzy set discusses the fuzziness degree of fuzzy set and was first mentioned in 1965 by Zadeh. Several scholars have studied it from different point of view. For example, De Luca and Termini [5] introduced some axioms to describe the fuzziness degree of fuzzy set. Another way given by Yager [6] was to view

^a Department of Mathematics, Science College, Kokrajhar, Assam, India

* Corresponding Author: Email: mamonidhar@gmail.com;

the fuzziness degree of fuzzy set in terms of lack of distinction between fuzzy set and its complement. Aimed at these two concepts, Kosko investigated fuzzy entropy in relation to subsethood measure. Kosko [2] has defined the entropy of a fuzzy set A as:

$$E(A) = \frac{M(A \cap A^c)}{M(A \cup A^c)} \quad (5)$$

Where $M(A \cap A^c)$ and $M(A \cup A^c)$ denote the cardinalities of the sets $A \cap A^c$ and $A \cup A^c$.

To relate subsethood measure with fuzzy entropy, Kosko proposes the following expression, given a subsethood measure c ; the entropy e generated by c is defined as

It can be put in form

$$e(A) = c(A \cup A^c, A \cap A^c) \quad (6)$$

Further, Young [3] sets out her axiomatization imposing that the measure

$$E(A) = S(A \cup A^c, A \cap A^c) \quad (7)$$

will be a fuzzy entropy.

It is important to mention here that all the results discussed above were obtained with the help of the existing definition of complementation of fuzzy sets which is defined in the following manner:

The membership function of a the complement of a fuzzy set A is defined as

$$\mu_{A^c}(x) = 1 - \mu_A(x), x \in X \quad (8)$$

In this article, our main intention is to revisit these two definitions from the standpoint of the new definition of complementation and thereafter we would like to put forward some suggestions in this regard. But before proceeding further with our work, we would like to discuss in brief about the new definition of complementation of fuzzy sets as introduced by Baruah [7, 8].

2. New Definition of Complementation of Fuzzy Sets

The existence of two functions of which one is called membership function and the other is reference function are required to define a fuzzy set, Baruah [7, 8]. The membership value is the difference between the membership function and reference function with the condition that the membership function should be greater than the reference function.

Accordingly,

A fuzzy set

$$A = \{x, \mu(x), x \in X\} \quad (9)$$

would be defined in this way as

$$A = \{x, \mu(x), 0, x \in X\} \quad (10)$$

so that the complement would become

$$A^c = \{x, 1, \mu(x), x \in X\} \quad (11)$$

It is important to mention here the fact that these definitions would give us the following results

$A \cup A^c = \text{the Universal set}$

And

$A \cap A^c = \text{the null set}$

which violates Zadeh's initial conception that fuzzy sets don't obey excluded middle laws.

Again it is important to note here that according to our definition $A \cap A^c$ is a null set and $A \cup A^c$ is the universal set and since a null set is always a subset of the universal set then we have

$$A \cap A^c \subseteq A \cup A^c \quad (12)$$

The above result discards Kosko's claim that the universal set can be a subset of any of its subset to some degree.

This concept of complementation of fuzzy sets plays a key role in developing some requirements of subsethood measure of fuzzy sets and thereby effects the relationship between entropy and subsethood established so far.

3. Comment on The Entropy Subsethood Relationship

From the above analysis we think that the better form of Young's definition may be the following:

A real function $c: F(X) \times F(X) \rightarrow [0,1]$ is called a subsethood measure, if c has the following properties

$$(C1) c(A,B)=1, \text{ if and only if } \mu_A(x) \leq \mu_B(x) \text{ i.e}$$

the membership values of the fuzzy set A is less than or equal to the membership values of the fuzzy set B.

$$(C2) c(A, A^c) = 0, \forall x \in X$$

$$(C3) c(A \cup A^c, A \cap A^c) = 0, \forall x \in X$$

$$(C4) \text{If } A \subseteq B \subseteq C \text{ then } c(C, A) \leq c(B, A) \text{ and if } A \subseteq B,$$

$$\text{then } c(C, A) \leq c(C, B)$$

It is important to mention here that the second condition of Young's definition is replaced by a new one and another additional condition is added to make it logical. The reason behind such a proposal can be described by the fact that the membership value of a fuzzy set can never remain included in the membership value of its complement set. The property (C2) comes in this way. Regarding (C3) we have discussed in the previous section.

If these conditions are taken for granted for subsethood then the entropy subsethood relationship proposed by the author would give us no idea of fuzziness of the fuzzy set concerned.

As regards Kosko's entropy subsethood relationship, we would like to draw attention to the following few lines.

By this formula it was derived that $A \cup A^c$ is a subset of $A \cap A^c$ to some degree and considered this as the unique feature of fuzzy set theory. Thus we can see that the Universe of discourse can also become a subset of any of its own subset to some degree. This is not desirable. Furthermore, the measure of fuzziness which was introduced by the author is also not free from defects as has been discussed in Dhar [12-15].

Another reason for which we would like to discard the subsethood theorem is due to the fact that. It was derived with the help of geometrical representation of fuzzy sets which is again controversial from the standpoints of the new definition of complementation on the basis of reference function. Let us have a look at it in brief in the following

Consider a universe of discourse containing two elements, $U = \{x_1, x_2\}$ The Universal set is represented by the point (1, 1) with the membership function $\mu_A(x_1) = 1$ and $\mu_A(x_2) = 1$. The point (1, 0) represents the set $\{x_1\}$ and the point (0, 1) represents the set $\{x_2\}$ Similarly, a fuzzy set defined in that universe of discourse $A = \{(x_1, \frac{1}{5}), (x_2, \frac{3}{5})\}$ is represented by the membership function

$$\mu_A(x_1) = \frac{1}{5} \text{ and } \mu_A(x_2) = \frac{3}{5}$$

Then the representation takes the following form

Then the point A can be represented as a point in two dimensional unit hypercube which is a square. This square represents all possible fuzzy sets of both elements; vertices of the square

represent crisp set.

The idea that the fuzzy set located at the vertex has the zero entropy and the fuzzy set located at the midpoint has the maximum entropy, led Kosko to define fuzzy entropy in the following form:

$$E(A) = \frac{M(A \cap A^c)}{M(A \cup A^c)} \quad (13)$$

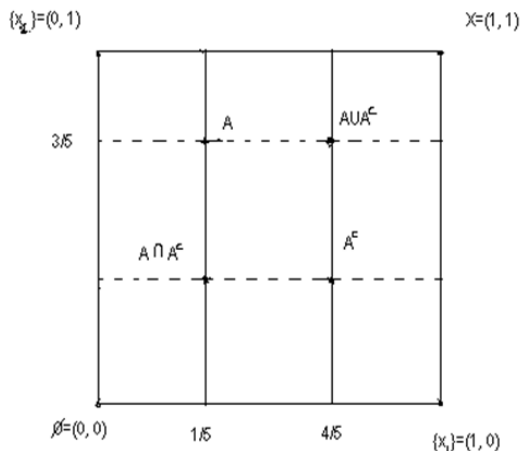


Figure 1. 2- Geometrical representation of fuzzy sets

where A^c stands for the complement of the fuzzy set A and $M(A)$ stands for the cardinality of the fuzzy set A .

The geometrical interpretation of fuzzy entropy theorem which is the outcome of geometrical representation of fuzzy sets was presented in the following form:

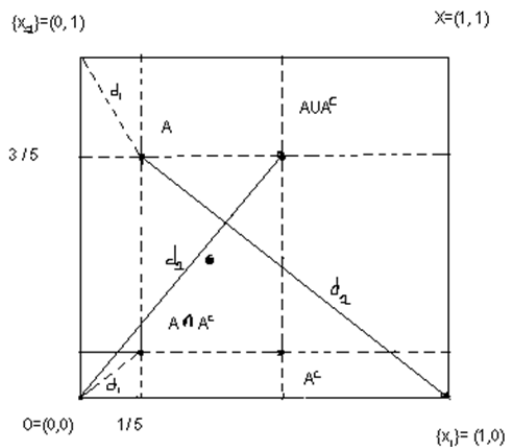


Figure 2. Geometrical representation of fuzzy entropy theorem

In the figure, d_1 represents $M(A \cap A^c)$ and d_2 represents $M(A \cup A^c)$

From the standpoint of the new definition of complementation $A \cap A^c$ coincides with the corner point \emptyset and $A \cup A^c$ coincides with the other corner point X having cardinalities zero and $|X|$ respectively. In order to get rid of such types of controversies, here we would like to suggest some properties of subsethood measure of fuzzy sets which is based on membership value.

The concept of new definition of complementation of fuzzy sets has led us to redefine the properties of subsethood of fuzzy sets when expressed in set theoretic form in the following way

$$(E1) 0 \leq S(A, B) \leq 1$$

(E2) $S(A, B) = 1$ if and only if $\mu_A(x) \leq \mu_B(x)$ i.e the membership values of the fuzzy set A is less than or equal to the membership values of the fuzzy set B .

$$(E3) S(A, A^c) = 0, \forall x \in X$$

$$(E4) S(A \cup A^c, A \cap A^c) = 0, \forall x \in X$$

$$(E5) \text{If } A \subseteq B \subseteq C \text{ then } S(C, A) \leq S(B, A) \text{ and if } A \subseteq B,$$

then $S(C, A) \leq S(C, B)$. At first blush, inclusion measure and entropy of fuzzy sets don't seem to be related. However with respect to specific pair of entropy and inclusion measure of fuzzy sets Kosko [2] and Young showed the result

$$E(A) = S(A \cup A^c, A \cap A^c) \quad (14)$$

But from the above discussion it is clear that the degree of subsethood of $A \cup A^c$ in $A \cap A^c$ is zero and hence we shall get no result from the said entropy subsethood relationship established so far.

4. Conclusions

In this article, we have commented on the two most popular entropy subsethood relationships and have given some suggestions. As regard subsethood some new properties are proposed which in turn discard the entropy-subsethood relationship.

References

- [1] Zadeh L A, Fuzzy Sets, Inform. and Control, 1965,8: 338-353.
- [2] B. Kosko, Fuzziness Vs Probability, Int.J.General Systems,17(1990),211-240
- [3] V.R.Young, Fuzzy Subsethood, Fuzzy Sets and Systems, 77(1996), 371-384.
- [4] D. Sinha and E.R. Dougherty, Fuzzification of set inclusion: theory and applications, Fuzzy Sets and Systems, 55(1993), 15-42
- [5] De Luca A, Termini S, A definition of non-probabilistic entropy in the settings of fuzzy set theory, Information and Control, 1972, 20:301-312.
- [6] Yager R.R, A procedure for ordering fuzzy subsets of the unit interval, Information Science, 24(1981), 143-161.
- [7] Baruah H K, Towards Forming a Field of Fuzzy Sets, International Journal of Energy Information and Communications, 2011, 2(1): 16 – 20.
- [8] Baruah H K, Theory of Fuzzy sets Beliefs and Realities, International Journal of Energy, Information and Communications, 2011, 2(2): 1-22.
- [9] L.Kitainik, Fuzzy inclusions and fuzzydictonomous decision procedures in Optimization Models Using Fuzzy Sets and Possibility, eds, J. Kacprzyk and S.Orlovski (Reidel, Dordrecht, 1987),pp-154-170
- [10] Dubious and Prade, Fuzzy sets and systems: Theory and Applications, Academic Press, New York(1980)
- [11] W. Bandler and L.Kohout, Fuzzy power set and fuzzy implication operators, Fuzzy Sets and Systems, 4(1980), 13-30.
- [12] Dhar M, On Hwang and Yang's definition of Entropy of Fuzzy sets, International Journal of Latest Trend Computing, 2011, 2(4): 496-497.
- [13] Dhar M, A Note on existing Definition of Fuzzy Entropy, International Journal of Energy Information and Communications, 2012, 3(1): 17-21.
- [14] Dhar M, On Separation Index of Fuzzy Sets, International Journal of Mathematical Archives, 2012, .3(3): 932-934.
- [15] Dhar M, On Geometrical Representation of Fuzzy Numbers, International Journal of Energy Information and Communications, 2012, 3(2): 29-34.