Jestr

Journal of Engineering Science and Technology Review 6 (2) (2013) 78-84

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Research Article

Disease-Related Gene Identification by an Improved Type-2 Fuzzy Method on Microarrays

Danling Wang[1] and Yanfei Wang[2,*]

[1]*Dept. of Mathematics and Physics, University of Science and Technology Beijing, Beijing,100083 - China*
[2]*Dept. of Sciences, China Agriculture University, Beijing,100083- China*

Received 15 May 2013; Accepted 25 July 2013

Abstract

Type-2 fuzzy set can control the uncertainty information more effectively than conventional type-1 fuzzy set because its membership function is three-dimensional. In this paper we perform the identification of disease-related genes based on DNA microarray by type-2 fuzzy set theory and established type-2 fuzzy membership function to describe the differences of the gene expression values generated form normal people's genes and patient's. We also make a comparison between our type-2 fuzzy method with the traditional fuzzy method. The results suggest the type-2 FM-d values are significantly different, which make the identifications more reasonable and convincing than the original one.

Keywords: Type-2 fuzzy set, Disease related gene, Microarray

1. Introduction

A DNA microarray is a multiplex technology applied in molecular biology. It consists of arrayed series of thousands of microscopic spots of DNA oligonucleotieds, called features, each containing picomoles of a specific DNA sequence known as probes or reporters. Since an array can contain tens of thousands of probes, a microarray experiment can accomplish many genetic tests in parallel. Therefore, microarray has dramatically accelerated many types of investigation [1]. Microarray is considered as an important tool for advancing the understanding of the DNA information, molecular, mechanism, biology and pathophysiology of critical illness. The expression of thousands of genes can be assessed, complex pathways can be more fully evaluated in a single experiment. Thus, microarrays could lead to discovery of new gene involved in disease processes.

Disease-associated gene identification is one of the most important areas of medical research today. It is known that certain diseases, such as cancer, are reflected in the change of the expression values of certain genes. For instance, due to genetic mutations, normal cells may become cancerous. These changes can affect the expression level of genes. Gene expression is the process of transcribing a gene's DNA sequence into RNA. A gene's expression level indicates the approximate number of copies of that gene's RNA produced in a cell and it is correlated with the amount of the corresponding proteins made. Analyzing gene expression data can indicate the genes which are differentially expressed in the diseased tissues. Several important breakthroughs and progress have been made [2].

One effective approach of identifying genes that are associated with a disease is to measure the divergence of two sets of values of gene expression. Usually, they are patients' and normal people's expression data. In order to identify the genes that are associated with disease, one need to determine from each gene whether or not the two sets of expression values are significantly different from each other (Liang et al., 2006). The two most popular methods to measure the divergence of two sets of values are t-test and Wilcoxon rank sum test [3]. According to Liang et al.'s work, both of these two methods have some limitations. The limitation of t-test is that it cannot distinguish two sets with close means even though the two sets are significantly different from each other. Another limitation is that it is very sensitive to extreme values. Although rank sum test overcomes the limitation of t-test in sensitivity to extreme values, it is not sensitive to absolute values. This might be advantageous to some application but not to others [2]. The word "different" itself is a fuzzy concept and fuzzy theory has many advantages in dealing with data containing uncertainty; therefore fuzzy approaches have been taken into consideration to analyze DNA microarrays. Liang et al. proposed the FM test and the experiment results suggests fuzzy method perform well in some data analyzing. However, some limitations still exist. The most obvious one is when the values of gene microarray data are very similar and lack over-expression, in which case the FM-d valves are very close or even equal to each other. That made the FM-test inadequate in distinguishing disease genes.

To overcome these problems, we introduce type-2 fuzzy set theory into the research of disease-associated gene identification. Type-2 fuzzy set is an extension of traditional fuzzy set, introduced by Zadeh [4]. Of course, employment of type-2 fuzzy sets usually increases the computational

* E-mail address: yfmu@sina.com
ISSN: 1791-2377

complexity in comparison with type-1 fuzzy sets due to the additional dimension of having to compute secondary grades for each primary membership. However, if type-1 fuzzy sets would not produce satisfactory results, employment of type-2 fuzzy sets for managing uncertainty may allow us to obtain desirable results [5]. Mizumoto and Tanaka have studied the set theoretic operations of type-2 sets, properties of membership grades of such sets, and have examined the operations of their algebraic product and algebraic sum [6]. Dubois and Prade have discussed the join and meet operations between fuzzy numbers under minimum t-norm [7]. Karnik and Mendel have provided a general formula for the extended sup-star composition of type-2 relations [8] [9]. Type-2 fuzzy sets have already been used in a number of applications, including decision making [10] [11], solving fuzzy relation equations [12], and pre-processing of data [13].

In this paper we establish the type-2 fuzzy membership function for identification of disease-associated genes on microarray data of patients and normal people. We call it type-2 fuzzy membership test (type-2 FM-test) and apply it to diabetes and lung cancer data. For the ten best-ranked genes of diabetes identified by the type-2 FM-test, 7 of them have been confirmed as diabetes associated genes according to genes description information in Genebank and the published literature. One more gene than original approaches is identified. Within the 10 best ranked genes identified in lung cancer data, 8 of them are confirmed by the literature which is associated with lung cancer treatment. The type-2 FM-d values are significantly different, which makes the identifications more reasonable and convincing than the original FM-test. In the next section, we introduce the theoretical background needed for a description of the type-2 FM-test which is detailed in Section Ⅲ, and we will give our results in Section Ⅳ.

2. Theoretical Background

2.1 Type-2 Fuzzy set

Type-2 fuzzy set was introduced as an extension of a type-1 fuzzy set by Zadeh [8]. Mizumoto and Tanaka studied the set theoretic operations and properties of membership grades of such sets and examined their operations of algebraic product and algebraic sum [6]. Karnik and Mendel extended the algorithms for performing union, intersection and complement for type-2 fuzzy sets [8]. Type-2 fuzzy sets and tools have now been widely used in many fields.

In this paper, we use the following notation and terminology: a type-2 fuzzy set in the universal set X, denoted as $\tilde{A}$, can be characterized by a type-2 fuzzy membership function $\mu_{\tilde{A}}$ (x, u) as

$$\tilde{A} = \int_{x\in X} \mu_{\tilde{A}}(x)/x = \int_{x\in X}\left[\int_{u\in J_x} f_x(u)/u\right]\Big/x,\ J_x \subseteq [0,1] \qquad (1)$$

where $\mu_{\tilde{A}}(x)$ is the membership grade of $x\in X$, which is a type-1 fuzzy set in [0, 1]; J_x is the primary membership of x, a memberships of the primary membership of x is called a secondary membership of x in $\tilde{A}$, denoted as $f_x(u)$. Figure 1 shows an example of a type-2 fuzzy membership function. The uncertainty in the primary membership, consisting of a bounded region by upper and lower membership functions is called the footprint of uncertainty (FOU) which can be expressed as

$$\mathrm{FOU}(\tilde{A}) = \bigcup_{\forall x\in X} J_x = \{(x,u): u\in J_x \subseteq [0,1]\}. \qquad (2)$$

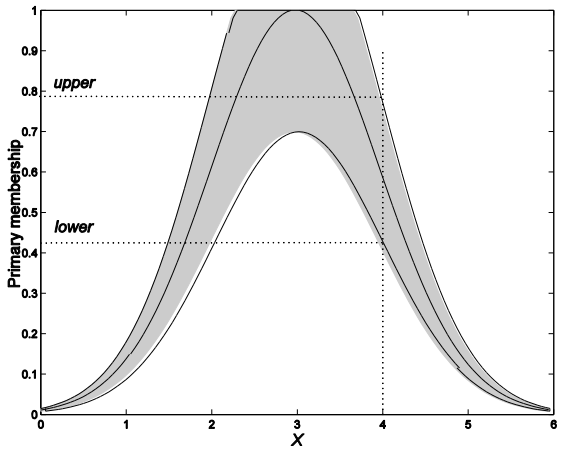

Fig. 1. Illustration of a type-2 fuzzy membership function

Although type-2 fuzzy sets can be useful in modeling uncertainty where type-1 fuzzy sets cannot, the operations of type-2 fuzzy sets involve numerous embedded type-2 fuzzy sets which include all possible combinations of secondary membership values. Therefore, a large number of computations are required. However, interval type-2 fuzzy sets can reduce the computational complexity significantly [15]. Due to this property, we apply the interval type-2 fuzzy set instead of a general type-2 fuzzy set in this paper. An interval type-2 fuzzy set is a specific type-2 fuzzy set when $f_x(u) = 1, \forall\, u\in J_x \subseteq [0, 1]$, and is expressed as

$$\tilde{A} = \int_{x\in X}\left[\int_{u\in J_x} 1/u\right]\Big/x. \qquad (3)$$

Figure 2 shows the secondary membership function of an interval type-2 fuzzy set at $x = 4$, in which the secondary memberships are all equally weighted for each primary membership of $x = 4$. Therefore, Jx can be expressed as

$$J_x = \{(x,u): u\in [\underline{u}_{\tilde{A}}(x), \overline{u}_{\tilde{A}}(x)]\} \qquad (4)$$

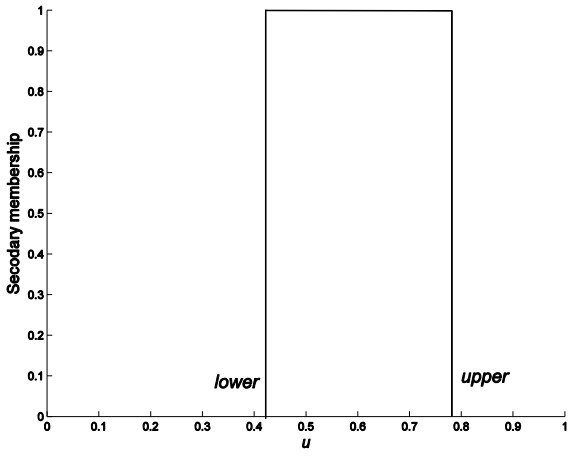

Fig. 2. Interval secodary membersip function

Where $\underline{u}_{\tilde{A}}(x)$ denote as lower bound of FOU and $\overline{u}_{\tilde{A}}(x)$ denote as upper bound of FOU. Moreover, FOU in (2) can be expressed as

$$\mathrm{FOU}(\tilde{A})=\bigcup_{\forall x\in X}[\underline{u}_{\tilde{A}}(x),\overline{u}_{\tilde{A}}(x)]\}$$

Figure 3 is a 3D plot of a type-2 fuzzy set whose secondary membership is a gauss function. As shown in the figure, the membership function of type-2 fuzzy sets is three-dimensional.

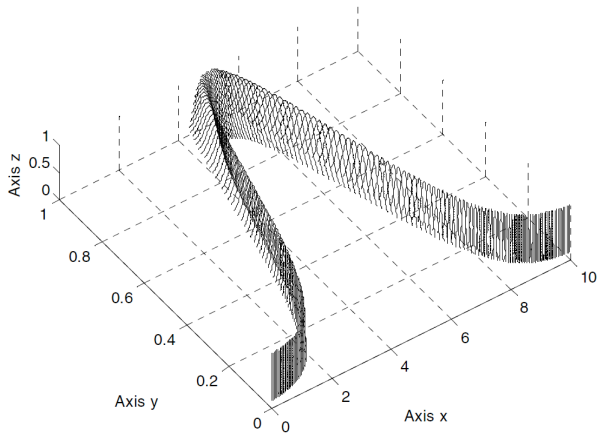

Fig. 3. 3D illustration of type-2 fuzzy set.

2.2 Type-2 Fuzzy Membership

In this paper we apply interval Gaussian type-2 fuzzy sets to identification of disease-related genes. The FOU can be described in terms of upper and lower membership functions. In the application we use upper and lower membership functions to establish primary membership functions of diabetes data and lung cancer data. For the Gaussian primary membership function with uncertain mean, the upper membership function $\overline{\mu}_{\tilde{A}}(x)$ is

$$\overline{\mu}_{\tilde{A}}(x)=\begin{cases}\mathcal{N}(m_1,\sigma;x) & x<m_1\\ 1 & m_1\le x\le m_2\\ \mathcal{N}(m_1,\sigma;x) & x>m_2\end{cases}$$

where, for instance, $N(m_1,\sigma;x)\equiv\exp[-\frac{1}{2}(\frac{x-m_1}{\sigma})^2]$.

The lower membership function $\underline{\mu}_{\tilde{A}}(x)$, is

$$\underline{\mu}_{\tilde{A}}(x)=\begin{cases}\mathcal{N}(m_2,\sigma;x) & x\le\dfrac{m_1+m_2}{2}\\ \mathcal{N}(m_1,\sigma;x) & x>\dfrac{m_1+m_2}{2}\end{cases}$$

For the Gaussian primary membership function with uncertain standard deviation, the upper membership function, $\overline{\mu}_{\tilde{A}}(x)$, is

$$\overline{\mu}_{\tilde{A}}(x)=\mathcal{N}(m,\sigma_2;x),$$

and the lower membership function, $\underline{\mu}_{\tilde{A}}(x)$, is

$$\underline{\mu}_{\tilde{A}}(x)=\mathcal{N}(m,\sigma_1;x).$$

These two examples illustrate how to define the upper and lower membership functions so that it is clear how to define them for other situations. However, for the problem in this paper, the upper and lower membership functions we established contain uncertainty both in mean and standard deviation. The plot is close to Figure 1.

2.3 Type Reduction

In a type-1 fuzzy logic system, the output corresponding to each rule is a type-1 set in the output space. To obtain the final result, defuzzification is needed. For type-2 fuzzy logic system, this step is considered as "type reduction" [15]. According to type reduction, a type-2 fuzzy set turns into a traditional fuzzy set and then output a crisp number to represent the combined output e.g., the centroid defuzzifier finds the union of all the out put sets and uses the centroid of the union as the crisp output. The traditional method to obtain centroid of a type-1 set A, whose domain is discretized into N points is given as

$$C_A=\frac{\sum_{i=1}^{N}x_i\mu_A(x_i)}{\sum_{i=1}^{N}\mu_A(x_i)}$$

3. Method

In this section, based on the FM-test of Liang's work we propose type-2 fuzzy membership test for disease-associated gene identification. We also consider S1 and S2 as two sets of values of a particular feature for two groups of samples under two different conditions, but this time we will establish type-2 fuzzy membership for the two sets $\tilde{S}_1$ and $\tilde{S}_2$. We choose the Gaussian function as the primary membership function. To avoid computational complexity, we apply the interval secondary membership function for this problem, which means all the secondary membership values are 1. Following the theoretical basis we introduced above, we should establish the upper and lower primary membership functions to describe the uncertainty in the gene expression data. In particular, this method is performed as follows:

1. Use the Gaussian function as the primary membership function to compute the mean (μ_1, μ_2) and standard deviation (σ_1, σ_2) of S_1 and S_2.

2. For each set, we establish the upper and lower primary membership functions. Here, both the mean and the standard deviation will be uncertain. We use two parameters

α and β which are in [0, 1] to control the uncertainty in mean and standard deviation respectively. In this paper, we set them all 0.5. Based on the FM-test and the rules of establishing upper and lower primary memberships for $\tilde{S}_1$, we obtain the upper primary membership as

$$\bar{\mu}_{\tilde{S}_1}(x)=\begin{cases} e^{-[x-(1-\alpha)\mu_1]^2/2(1+\beta)\sigma_1^2} & x<(1-\alpha)\mu_1 \\ 1 & (1-\alpha)\mu_1 \le x \le (1+\alpha)\mu_1 \\ e^{-[x-(1+\alpha)\mu_1]^2/2(1+\beta)\sigma_1^2} & x>(1+\alpha)\mu_1 \end{cases},$$

and the lower primary membership as

$$\underline{\mu}_{\tilde{S}_1}(x)=\begin{cases} e^{-[x-(1+\alpha)\mu_1]^2/2(1-\beta)\sigma_1^2} & x \le \mu_1 \\ e^{-[x-(1-\alpha)\mu_1]^2/2(1-\beta)\sigma_1^2} & x>\mu_1 \end{cases},$$

We can obtain the upper and lower primary membership functions similarly for $\tilde{S}_2$:

$$\bar{\mu}_{\tilde{S}_2}(x)=\begin{cases} e^{-[x-(1-\alpha)\mu_2]^2/2(1+\beta)\sigma_2^2} & x<(1-\alpha)\mu_2 \\ 1 & (1-\alpha)\mu_2 \le x \le (1+\alpha)\mu_2 \\ e^{-[x-(1+\alpha)\mu_2]^2/2(1+\beta)\sigma_2^2} & x>(1+\alpha)\mu_2 \end{cases},$$

$$\underline{\mu}_{\tilde{S}_2}(x)=\begin{cases} e^{-[x-(1+\alpha)\mu_2]^2/2(1-\beta)\sigma_2^2} & x \le \mu_2 \\ e^{-[x-(1-\alpha)\mu_2]^2/2(1-\beta)\sigma_2^2} & x>\mu_2 \end{cases},$$

3. Use the upper and lower primary membership functions $\bar{\mu}_{\tilde{S}_i}(x)$ and $\underline{\mu}_{\tilde{S}_i}(x)$, i = 1, 2; and the secondary membership values fx(u) to quantify the convergence of S_1 and S_2. Type-reduction work is needed in this step. Here, since we use the interval type-2 fuzzy set, fx(u) =1, $\forall u \in J_x \subseteq [0,1]$. The secondary memberships are all uniformly weighted for each primary membership of x.

4.Calculate the divergence degree between the two sets based on the convergence degree.

Type-reduction is an important step for type-2 fuzzy sets. In our application, $\forall x \in X$, a primary membership interval $[\underline{\mu}_{\tilde{S}_i}(x), \bar{\mu}_{\tilde{S}_i}(x)]$ can be obtained. We discretize it into N points, where $a_1=\underline{\mu}_{\tilde{S}_i}(x)$ and $a_N=\bar{\mu}_{\tilde{S}_i}(x)$; then the final membership of x can be obtained as

$$\mu(x)=\frac{\sum_{i=1}^{N} a_i \times f_x(a_i)}{N}.$$

This type reduced membership $\mu(x)$ maps each value x in S_1 or S_2 into a membership value to quantify the degree that x belongs to type-2 fuzzy set $\tilde{S}_1$ or $\tilde{S}_2$. For simplicity, we put $a_i = (a_{i-1} + a_{i+1}) / 2$, i = 2, …, N-1, while $f_x(a_i)$ = 1, $x \in X$, i = 1,…, N. can be expressed as

$$\mu(x)=\frac{\underline{\mu}_{\tilde{S}_i}(x)+\bar{\mu}_{\tilde{S}_i}(x)}{2},$$

To compute the overall bond between S_1 and S_2, we define type-2 FM c-values and d-values based on Liang et al. [2].

Definition 3.13 (Type-2 FM c-values): Given two sets S1 and S2, the convergence degree between S1 and S2 in FM-test is defined as

$$c(S_1,S_2)=\frac{\sum_{x\in S_1}\mu_{\tilde{S}_2}(x)+\sum_{y\in S_2}\mu_{\tilde{S}_1}(y)}{|S_1|+|S_2|}.$$

We define the divergence value as follows:

Definition 3.14: Given two sets S_1 and S_2, the divergence degree between S1 and S2 in the FM-test is defined as

$$d(S_1,S_2)=1-c(S_1,S_2)$$

4. Result and Discussion

In this section, we apply type-2 FM-test to a diabetes expression dataset and a lung cancer expression dataset, respectively. Meanwhile, we make a comparison with the results of traditional FM-test by Liang et al. The first dataset is a diabetes dataset of microarray gene expression data. It contains 10831 genes and is downloaded from [16]. For each gene in this dataset, there are 10 expression values, five from a group of insulin-sensitive (IS) people and five from a group of insulin-resistant (IR) people. To make this data more reliable, only the genes that have null expression values are included in this analysis. Meanwhile, we also require that, for a gene to be included, at least five out of its ten expression values are greater than 100.

4.1 Analysis of Diabetes Dataset

Ten best-ranked genes of diabetes identified by the type-2 FM-test and the original FM-test are shown in Table I. From this table we see that the results of the two methods are not much different. We identified gene U61734 and gene Z26491

by type-2 FM-test instead of gene U06452 and gene X57959 identified by the FM-test. Gene U06452 and gene X57959 are just candidate genes for diabetes mentioned in some published papers, but no biological evidence was put forward to confirm that they are related with diabetes [2, 16]. However, for gene Z26491, the homo sapiens gene for catechol o-methyltrans-fease (COMT), was found to be differently expressed and helpful for treatment in diabetic rats. Ref. [17] compared the activity of COMT in the livers of diabetic rats with that in normal rats; the result suggested the activity of COMT is lower in diabetic rats than in normal rats. Lal et al [18] examined the effect of nitecapone, an inhibitor of the dopamine-metabolizing enzyme COMT and a potent antioxidant, on functional and cellular determinants

TABLE I

Type-2 FM-test		
Prob Set	Gene Description	T2 d-value
U49573	Human phosphatidylinositol (4,5) bisphosphate	0.9867
U61734	Homo sapiens transmembrane emp24-like trafficking protein 10	0.9476
Z26491	Homo sapiens gene for catechol o-methyltrans-fease	0.9126
X81003	Homo sapiens HCG V mRNA	0.8732
L07033	Human hydroxymethylglutaryl-CoA lyase mRNA	0.8471
D85181	Homo sapiens mRNA for fungal sterol-C5-desaturase homolog	0.8426
L07648	Human MXII mRNA	0.8127
X53586	Human. mRNA for integrin alpha 6	0.7851
M60858	Human. nucleolin gene	0.7662
M95610	Human alpha 2 type IX collagen (COL9A2) mRNA	0.7565
FM-test		
Probe Set	Gene Description	FM d-value
U45973	Human phosphatidylinostiol (4,5) bisphosphate	0.9988
D85181	Homo sapiens mRNA for fungal sterol-C5-desaturase homolog	0.9351
M60858	Human nucleolin gene	0.8918
M95610	Huamn alpha 2 type IX collagen (COL9A2) mRNA	0.8718
X53586	Human mRNA for integrin alpha 6	0.8513
L07648	Human MXII mRNA	0.8575
L07033	Human hydroxymethylglutaryl-CoA lyase mRNA	0.8554
X81003	Homo sapiens HCG V mRNA	0.7914
X57959	Ribosomal protein L7	0.7676
U06452	Melan-A	0.7566

TABLE II

Type-2 FM-test		
Prob Set	Gene Description	T2 d-value
NM_002405	MFNG: MFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase	0.8423
NM_001335	CTSW: cathepsin W	0.8164
NM_0002694	ALDH3B1: aldehyde dehydrogenase 3 family, member B1	0.7769
NM_024830	LPCAT1: lysophosphatidylcholine acyltransferase 1	0.7532
AA888858	PDE3B: phosphodiesterase 3B, cGMP-inhibited	0.7247
BE789881	RAB31: member RAS oncogene family	0.7211
NM_006079	CITED2: cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain,2	0.7062
AF026219	DLC1:deleted in liver cancer 1	0.6985
AV728526	DTX4: deltex homolog 4 (Drosophila)	0.6932
NM_004415	DSP: Desmoplakin	0.6713
FM-test		
Probe Set	Gene Description	FM d-value
NM_173086	KRT6E: Keratin 6E	1
NM_001723	DST: Dystonin	1
NM_002639	SERPINB5: Serpin peptidase inhibitor, clade B (ovalbumin), member 5	1
AB010153	TP73L: Tumor protein p73 like	1
NM_023915	GPR87: G protein-coupled receptor 87	1
NM_006536	CLCA2: Chloride channel, calcium activated, family member 2	1
NM_00100533 7	PKPI: Plakophilin 1 (ectodermal dysplasia/skin fragility syndrome)	1
AF043977	CLCA2: Chloride channel, calcium activated, family member 2	1
NM_019093	UGTIA9: UDP glucuronosyltransferase ! family, polypeptide A9	1
NM_004415	DSP: Desmoplakin	1

of renal function in rats with diabetes. The results suggested that the COMT inhibitory and antioxidant properties of nitecapone provide a protective therapy against the development of diabetic nephropathy [18]. These works proved that gene Z26491 is related with diabetes or treatment. In the remaining genes, U45973, M60858, L07648, L07033, X53586, X81003 are diabetes associated genes according to the current literature [2]. D85181, M95610 and U06452 are recommended as candidate diabetes genes by Liang et al.[2], while X57959 was recommended in Yang et al. [12]. We recommended gene U61734 as candidate gene for future research in this field. From the comparison, one more disease-associated gene is identified by the type-2 FM-test.

4.2 Analysis of Lung Cancer Dataset

Table II shows ten best-ranked genes of lung cancer identified by the two approaches. The results are very different. In the result obtained by the type-2 FM-test, 8 disease associated genes are identified. MFNG is a member of the fringe gene family which also includes radical and lunatic fringe genes. They all encode evolutionarily conserved secreted proteins that act in the Notch receptor pathway. The activity of fringe proteins can alter Notch signaling [19]. Activation of the Notch 1 signaling pathway can impair small cell lung cancer viability [20]. The protein encoded by CTSW is found associated with the membrane inside the endoplasmic reticulum of natural killer (NK) [19]. NK cells play a major role in the rejection of tumors and cells infected by viruses [21]. ALDH3B1 is highly expressed in kidney and lung [19]. Marchitti et al. [22] found ALDH3B1 expression was upregulated in a high percentage of human tumors; particularly in lung cancer cell the value is highest. Increasing ALDH3B1 expression in tumor cells may confirm a growth advantage or be the result of an induction mechanism mediated by increasing oxidative

stress [22]. LPCAT1 activity is required to achieve the levels of SatPC essential for the transition to air breathing [23] and it is also upregulated in cancerous lung [24]. Gene PDE3B was mentioned in [25] as the most significantly amplified gene in the tumors. CITED2 is required for fetal lung maturation [26]. Researchers found CITED2 was highly expressed in lung cancer but not in normal tissues, which demonstrates that CITED2 plays a key role in lung cancer progression [27]. Gene DLC1 encodes protein deleted in liver cancer [2]. This gene is deleted in the primary tumor of hepatocellular carcinoma. It maps to 8p22-p21.3, a region frequently deleted in solid tumors. It is suggested that this gene is a tumor suppressor gene for human liver cancer, as well as for prostate, lung, colorectal and breast cancers [19].

8 disease-associated genes are identified by the original FM-test method [2]. However, when we applied the FM-test on lung cancer data, there are more than 80 genes having the same FM d-values; they are all equal to one, which made it difficult to rank and distinguish disease associated genes from others. We have to choose the overexpressed genes from these 80 genes for analysis, which made the task more complicated, and it may miss some important genes. The reason is that the gene expression values in lung cancer microarray data are very close to each other, and the original data is noisy. The FM-test does not seem to be able to deal with these factors suitably.

5. Conclusion

We proposed a type-2 FM-test approach based on the type-2 fuzzy set theory and the work of Liang *et al.* [2]. This approach was then applied on diabetes and lung cancer microarray data. For each data set, we analyzed 10 best-ranked genes. In diabetes data, we identified 7 genes which have been confirmed to be related to diabetes or treatment in the published literature. In lung cancer data, we also identified 7 genes. The type-2 FM d-values are significantly different for each gene, which makes the ranking more reasonable than from the original FM-test. The type-2 FM test performs better than the original FM test when analyzing microarray data containing similar expression values and noise.

6. Acknownledgment

This work was supported in part by the Chinese Universities Scientific Fund (Grant No. 2013XJ010) and the Chinese Fundamental Research Funds for the Central Universities (Grant No. FRF-TP-13-020A)

References

1. Maskos, U. and Southern E.M., “onglass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides sysnthesised in situ”, Nucleic Acides Res, Vol. 11, 1992, pp. 1679-1684.
2. Liang, L.R., Lu, S., Wang, X., Mandal, V., Patacsil, D. and Kumar, D., “FM-test: a fuzzy-set-theorybased approach to differential gene expression data analysis,” BMC Bioinformatics, 7 (4), 2006,pp: 54-69.
3. Rosner, B., Fundamentals of Biostatistics. In Pacific rove 5th edition, CA:Duxbury Press, 2000.
4. Zadeh, L.A., “The concept of a linguistic variable and its application to approximate reasoning-1” Inform. Sci., Vol. 8, 1975, 199-249.
5. Hwang, C. and Rhee, F.C.-H “Unvertain fuzzy clustering: Interval type-2 fuzzy approach to C-Means”, IEEE Transactions on Fuzzy Systems, Vol. 15, No.1, 2007, pp.107-120.
6. Mizumoto, M. and Tanaka, K., “Some properties of fuzzy sets of type-2 fuzzy sets”, Inform, Control, Vol. 31, 1976, pp. 312-340.
7. Dubois, D and Prade, H, 1980, “Fuzzy sets and systems: Theory and applications”, Academic press, INC, Chestnut Hill, MA.
8. Karnik, N.N. and Mendel, J.M., “Operations on type-2 fuzzy sets”, Int. J. Fuzzy sets Syst., Vol, 122, 2001, pp. 327-348.
9. Karnik, N.N. and Mendel, J.M., “Type-2 fuzzy logic systems”, IEEE Trans on Fuzzy Systems, Vol. 7, 1999, pp.643-658.
10. Chaneau, J.L., Gunaratne, M. and Altchaeffl, A.G., “An application of type-2 sets to decision making in engineering, in: J.C. Bezdek, Analysis of fuzzy information- vol. II: Artificial intelligence and Decision systems”, CRC Press, Boca Raton, FL. 1987.
11. Yager, R.R, “Fuzzy subsets of typeII in decisons”, J. Cybernet, Vol. 10, 1980, pp. 137-159.
12. Wagenknecht, M., Hartmann, K., “Application of fuzzy sets of type 2 to the solution of fuzzy equation systems”, Fuzzy Sets Systems, Vol. 25, 1988, pp. 183-190
13. John, R.I., “Type2 fuzzy sets: an appraisal of theory and applications”, Int, J.Uncertainty, Fuzzinewss Knoledge-bases systes, VOl.6, No.6, pp.563-576.
14. Mendel J.M. and Bob John, R.I., “Type-2 fuzzy sets made simple”, IEEE Trans. On fuzzy systems, Vol. 10, 2002, pp.117-128.
15. Mendel, J.M. “Uncertain rule-based fuzzy logic systems: introduction and new directions”, Prentice Hall PTR, Upper Saddle River, NJ. 2002.
16. Yang, X., Pratley, R.T, Tokraks, S., Bogardus, C., Permana, P.A.,”Microarray profiling of skeletal muscle tissues from equally obese, non-diabetic insulin-sensitive and insulin-resistant Pima Indians”, Diabetologia, vol. 45, 2002, pp1584-1593.
17. Wang, J., Liu, I, Tzeng, T and Chen, J., “Decrease in catechol-o-methytransferase activity in the liver of streptozotocin-induced, diabeticrats”, Clin Exp Pharmacol& Physiol, Vol. 29, 2002, pp.419-422.
18. Lal, M. A., Krner, Al, Matsuo, Y., Zelenin, S., Cheng,X., Jarmko, G., Dibona, G., Ekol, A, Aperia, A., “Combined Antioxdant and COMT inhibitor treatment reverses renal abnormalities in diabetic rats”, Diabets, Vol. 49, 2000, pp. 1381-1389.
19. http://www.ncbi.nlm.nih.gov/genbank/.
20. Platta, C.S., Greenblatt, D, Y., Junnimalaiyaan M. and Chen, H., “Valpproic acid induces Notch 1 signaling in small cell lung cancer cells”, J. Surgcal Research, Vol. 148, 2000, pp. 31-37.
21. Oldham, R, “Natural killer cells: Artifact to reality: an odyssey in biology”, Cancer Metastasis Reviews, Vol. 2, 1983, pp.323-326.
22. Marchitti, A., Orlicky, D.J., Brocker, C. and Vasiliou, V., “Aldehyde, Dehydrogenase 3B1 (ALD3B1): Immunohistochemical tissue distribution and cellular-specific localization in normal and cancerous human tissue” J. Histochemistry and Cytochemistry, Vol. 58, 2010, pp. 765-783.
23. Bridges, J.P. Ikegami, M. Brilli, L.L., Chen, X., Mason, R.J., Shannon, J.M., “LPCAT1 regulates surfactant phospholipids synthese and is required for trasitioning to air breathing in mice”, J. Clinical Investigation, vol. 120, 2010, 1736-1748.
24. Mansilla, F., Costa, K., Wang, S., Kruhoffer, M., Lewin, T.M., Orntoft, T.F., Coleman, R.A. and Birkenkmp-Demtroder, K., “Lysophosphatidylcholine acylransferase 1(LPCAT1) overexpression in human colorectal cancer”, J.Mol Med, Vol. 87, 2009.
25. Lo, K. C., Stein, L.C., Panzarella, J.A., Cowell, J.K., Hawthorn, L., “Identification of genes involved in squamous cell carcinoma of the lung using synchronized data from DNA copy number and transcript exrepssion profiling analysis”, Lung Cancer, Vol. 59, 2008, pp. 315-331.
26. Xu, B., Qu, X., Doughman, S., Watanabe, M., Dunwoodie, S.L., Yang ,Y., “Cited 2 is required for featal lung maturation”, Dev Biol, Vol. 317, 2008, pp.95-105.

27. Chou, Y.T., Hsu, C.F., Kao, Y.R. and Wu, C.W., "Cited 2, a novel EGFR-induced coactvator, plays a key role in lung cancer progession", AACR 101st Annual Meeting, Poster Presentation. 2010.