RATIO-HYPOTHESIS-BASED FUZZY FUSION WITH APPLICATION TO CLASSIFICATION OF CELLULAR MORPHOLOGIES

Tuan D. Pham
School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2600, Australia

Xiaobo Zhou
Center for Biotechnology and Informatics, Methodist Hospital Research Institute, Houston, TX 77030, U.S.A.

Keywords: Permanence of ratio, Information fusion, Pattern classification, Cellular phenotypes, Bio-imaging.

Abstract: Fusion of knowledge from multiple sources for pattern recognition has been an active area of research in many scientific disciplines. This paper presents a fuzzy version of a probabilistic fusion scheme, known as permanence-of-ratio-based combination, with application to analysis of cellular imaging for high-content screening. Classification of cellular phenotypes has been carried out to illustrate the usefulness of the permanence-of-ratio-based fuzzy fusion.

1 INTRODUCTION

Information or data fusion can be defined as the use of mathematical methods that combine data from multiple sources to obtain the resultant knowledge in order to achieve inferences, which will be more efficient and potentially more accurate than if they were achieved by means of considering separate single sources. Information fusion can be performed on a low-level or high-level process depending on the processing stage at which fusion takes place. Low-level fusion combines several sources of raw data to produce new raw data (Pellizzeri *et al.*, 2002). The expectation is that fused data is more informative and synthetic than the original raw data. High-level fusion typically combines features from multiple classifiers, or signals from multiple sensors for logical decision making (Muller *et al.*, 2001; Das, 2008).

There are many mathematical operators developed for data fusion such as the averaging rule, multiplication rule, probabilistic models, mathematical theory of evidence, machine learning methods, and fuzzy integral (Chi *et al.*, 1996). For high-level fusion, the rationale of combining knowledge from various sources is that it is always difficult or impossible to design a single classifier or to use a single feature for pattern classification to achieve the best results, because a particular classifier or feature can only be robust for handling a particular identity of an object, which may vary under different settings. Furthermore, different problems may require different data fusion methods to obtain effective solutions depending on the types of features.

This paper discusses the use of fuzzy measures for combining evidences from multiple sources, where the strong assumption of data independence is relaxed (Journel, 2002). The utilization of such novel idea appears to be promising for pattern classification but it is still rarely explored in the new field of bioinformatics. We are interested in applying an information fusion scheme for combining output from multiple classifiers in order to improve the results for classifying various cellular phenotypes for robust automated analysis of genome-wide high-content screening of fluorescent microscopy images of cells.

The rest of this paper is organized as follows. Section 2 introduces the concept of the permanence of ratio hypothesis for knowledge combination. Section 3 proposes a ratio-hypothesis-based fuzzy fusion scheme. Section 4 illustrates the application of the proposed fusion model for classifying cellular phenotypes for genome-wide screening, such screening is essential to the rapid discovery of basic biological cell principles such as control of cell cycle and cell mor-

D. Pham T. and Zhou X. (2010).
RATIO-HYPOTHESIS-BASED FUZZY FUSIONWITH APPLICATION TO CLASSIFICATION OF CELLULAR MORPHOLOGIES.
In *Proceedings of the Third International Conference on Bio-inspired Systems and Signal Processing*, pages 202-207
DOI: 10.5220/0002707902020207

phology. Section 5 presents the use of a Gaussian distribution for estimating fuzzy densities. Finally, the conclusion of the finding is given in Section 6.

2 INFORMATION FUSION USING PERMANENCE OF RATIO HYPOTHESIS

Based on the engineering paradigm of the permanence of updating ratios, which asserts that the rates or ratios of increments are more stable than the increments themselves, as an alternative to the assumption of the full or conditional independence of probabilistic models; Journel introduced a scheme for information fusion of diverse sources (Journel, 2002). This scheme allows the combination of data events without having to assume their independence. This information fusion is described as follows.

Let $P(A)$ be the prior probability of the occurrence of data event A; $P(A|B)$ and $P(A|C)$ be the probabilities of occurrence of event A given the knowledge of events B and C, respectively; $P(B|A)$ and $P(C|A)$ the probabilities of observing events B and C given A, respectively. Using Bayes' law, the posterior probability of A given B and C is

$$\begin{aligned} P(A|B,C) &= \frac{P(A,B,C)}{P(B,C)} \\ &= \frac{P(A)P(B|A)P(C|A,B)}{P(B,C)} \end{aligned} \quad (1)$$

The simplest way for computing the two probabilistic models is to assume the model independence, giving $P(C|A,B) = P(C|A)$, and $P(B,C) = P(B)P(C)$. Thus, (1) can be rewritten as

$$\frac{P(A|B,C)}{P(A)} = \frac{P(A|B)}{P(A)} \frac{P(A|C)}{P(A)} \quad (2)$$

However, the assumption of conditional independence between the data events usually does not statistically perform well and leads to inconsistencies in many real applications (Journel, 2002). Therefore, an alternative to the hypothesis of conventional data event independence should be considered. The permanence of ratios based approach allows data events B and C to be incrementally conditionally dependent and its fusion scheme gives

$$P(A|B,C) = \frac{1}{1+x} = \frac{a}{a+bc} \in [0,1] \quad (3)$$

where

$$a = \frac{1-P(A)}{P(A)},\ b = \frac{1-P(A|B)}{P(A|B)},$$
$$c = \frac{1-P(A|C)}{P(A|C)},\ x = \frac{1-P(A|B,C)}{P(A|B,C)}.$$

An interpretation of the fusion expressed in (3) is as follows. Let A is the target event which is to be updated by events B and C. The term a is considered as a measure of prior uncertainty about the target event A or a distance to the occurrence of A without any updated evidence. We have a = 0 for $P(A)$ = 1 if target event A is certain to occur; and a = 0 for $P(A)$ = 0 if A is an impossible event. Likewise, b and c measure the distances to A knowing about its occurrence after observing evidences given by B and C, respectively. The term x is the distance to the target event A occurring after observing evidences given by both events B and C. The ratio c/a is then the incremental (increasing or decreasing) information of C to that distance starting from the prior distance a. Similarly, the ration x/b is the incremental information of C starting from the distance b. Thus, the permanence of ratios provides the following relation

$$\frac{x}{b} \approx \frac{c}{a} \quad (4)$$

which says that the incremental information about C to the knowledge of A is the same after or before knowing B. In other words, the incremental contribution of information from C about A is independent of B. This expression relaxes the restriction of the assumption of full independence of B and C.

For the generation of k data events E_j, $j = 1,\ldots,k$; the conditional probability provided by a succession of $(k-1)$ permanence of ratios is given as

$$P(A|E_j,\ j = 1,\ldots,k) = \frac{1}{1+x} \in [0,1] \quad (5)$$

where

$$x = \frac{\prod_{j=1}^{k} d_j}{a^{k-1}} \geq 0$$

$$a = \frac{1-P(A)}{P(A)}$$

$$d_j = \frac{1-P(A|E_j)}{P(A|E_j)},\ j = 1,\ldots,k$$

It is clear that expression (5) requires only the knowledge of the prior probability $P(A)$, and the k elementary single conditional probabilities $P(A|E_j)$, $j = 1,\ldots,k$, which can be independently computed.

We next present the concept of fuzzy measures and how we can implement fuzzy measures into the framework of permanence-of-ratio hypothesis for information fusion, where the independence of joint events can be more relaxed to provide a better model for information consistency.

3 RATIO-HYPOTHESIS-BASED FUZZY INFORMATION FUSION

Let X be a finite set $X = \{x_1, x_2, \ldots, x_n\}$. A fuzzy measure g defined on X is a set function $g : \mathcal{P}(X) \to [0,1]$ satisfying the following axioms (Sugeno, 1977):

1. $g(\emptyset) = 0$, and $g(X) = 1$.
2. If $A \subseteq B$, then $g(A) \leq g(B)$.

where $\mathcal{P}(X)$ denotes the power set of X.

It is noted that when the second property is not satisfied, g is called a non-monotonic fuzzy measure (Grabisch, 1996). There are 2^n coefficients being equivalent to the cardinality of $\mathcal{P}(X)$ to compute a fuzzy measure on X. These coefficients are the values of g for all subsets of X and they are not independent since they must satisfy the property of monotonicity. Theoretically, the concept of fuzzy measures is the generalization of the classical measure theory which is restrictive on the hypothesis of additivity; whereas additivity is relaxed by the theory of fuzzy measures.

Sugeno (1977) defined a fuzzy measure known as the g_λ-fuzzy measure that satisfies the following additional condition, $\forall A, B \subset X$, and $A \cap B = \emptyset$,

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B),\ \lambda > -1. \tag{6}$$

To simplify the notation, let $g^i = g(\{x_i\})$ which is called a fuzzy density function. A fuzzy density g^i can be interpreted as the degree of belief or degree of importance that the corresponding attribute x_i makes an effect or contribution towards the whole fuzzy system when all attributes are considered together. Let $A = \{x_{i_1}, x_{i_2}, \ldots, x_{i_m}\} \subset X$, $g_\lambda(A)$, $\lambda \neq 0$, can be expressed as(Lesczynski *et al.*, 1985)

$$g_\lambda(A) \quad = \quad \frac{1}{\lambda}\left[\prod_{x_i \in A}(1 + \lambda g^i) - 1\right] \tag{7}$$

The value of λ can be calculated using the condition $g(X) = 1$ as follows.

$$\lambda + 1 = \prod_{i=1}^{n}(1 + \lambda g^i) \tag{8}$$

The following properties of the g_λ-fuzzy measure will be helpful in the computation of the parameter λ (Tahani and Keller, 1990).

1. Lemma: For a finite set $\{g^i\}$, $0 < g^i < 1$, there exists a unique root $\lambda \in (-1, +\infty)$, and $\lambda \neq 0$. Based on this lemma, λ can be determined by solving $(n-1)$ degree polynomial and selecting the unique root > -1.
2. If $\sum_{i=1}^{n} g^i < 1$, then $\lambda > 0$.
3. If $\sum_{i=1}^{n} g^i > 1$, then $-1 \leq \lambda < 0$.

Among other computer methods being useful for clinical applications such as Mycine (Shortlie, 1976) and several other medical expert systems (Berner, 1998), the Shafer's theory of evidence (Shafer, 1976) is a popular tool for medical decision making (Klir and Wierman, 1999). There are some connections between the belief and plausibility measures of the theory of evidence and the Sugeno's fuzzy measures. The function which maps $\mathcal{P}(X)$ to $[0,1]$ is called a belief function, denoted as *bel*, iff it satisfies the following conditions (Shafer, 1976):

1. $bel(\emptyset) = 0$, $bel(X) = 1$
2. $bel(\bigcup_i A_i) \geq \sum_{\emptyset \neq I \subseteq \{x_1, x_2, \ldots, x_n\}} (-1)^{|I|+1} bel(\bigcap_i A_i)$

The plausibility measure is defined in terms of the belief measure as

$$pl(A) = 1 - bel(\bar{A}) \tag{9}$$

Some other mathematical relationships between belief and plausibility measures, $\forall A \in \mathcal{P}(X)$, are

$$bel(A) + bel(\bar{A}) \leq 1,\ pl(A) + pl(\bar{A}) \geq 1$$

and

$$pl(A) \geq bel(A)$$

Based on the definitions and properties of the belief and plausibility measures, Banon (1981) has shown that a g_λ-fuzzy measure is a belief measure when $\lambda \geq 0$, and a g_λ-fuzzy measure is a plausibility measure when $\lambda \leq 0$.

By allowing the calculation of the joint fuzzy events $g(A|B,C)$ which makes the terms b and c defined in (5) equal to each other, we can equivalently define a fuzzy probabilistic fusion operator, denoted as $\mathcal{F}$, for a target event A which is to be updated by events B and C, as

$$\mathcal{F}(A|B,C) = \frac{a}{a+f} \tag{10}$$

where

$$a = \frac{1-g(A)}{g(A)}, \text{ and } f = \frac{1-g(A|B,C)}{g(A|B,C)}$$

The general form for the ratio-hypothesis-based fuzzy fusion can be defined by generalizing (10) to account for the fuzzy measure of k events $E_j, j = 1,\ldots,k$, giving

$$\mathcal{F}\,(A|E_j, j=1,\ldots,k) = \frac{a}{a+x} \in [0,1] \qquad (11)$$

where

$$x = \frac{1-g(A|E_1,\ldots,E_k)}{g(A|E_1,\ldots,E_k)}$$

Based on both (10) and (11), the ratio-hypothesis-based fuzzy fusion of multiple events allows a stronger conditional dependence than the ratio-hypothesis-based fusion expressed in (3) and (5) but still only requires the knowledge of the fuzzy densities of the multiple events. We next discuss how to estimate the fuzzy densities using the Gaussian probability function and also present the Bayes classifier.

4 BAYES CLASSIFIER AND ESTIMATING FUZZY DENSITIES BY GAUSSIAN FUNCTION

Pattern recognition using decision-theoretic framework is based on a discriminant or decision function to assign the unknown pattern to the best match. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ be an n-dimensional feature vector; and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ the set of m distinct patterns. The Bayes classifier for a 0-1 loss function is expressed as (Gonzalez and Woods, 2002)

$$d_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i);\ i = 1,\ldots,m. \qquad (12)$$

where $d_i(\mathbf{x})$ is a decision function that measures how likely the unknown pattern $\mathbf{x}$ belongs to the ith pattern class, $p(\mathbf{x}|\omega_i)$ is the probability density function of the feature vector of class ω_i, and $P(\omega_i)$ is the probability that class ω_i occurs.

The recognition procedure is to compute the m decision function $d_i(\mathbf{x})$, $i = 1,\ldots,m$; and then assign the pattern to the class whose decision function value is maximum. Using the Gaussian probability distribution function, its n-dimensional form is given as

$$p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{n/2}(\det \mathbf{C}_i)^{1/2}}\, e^{-\frac{1}{2}[(\mathbf{x}-\mathbf{m}_i)^T \mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i)]} \qquad (13)$$

where $\mathbf{C}_i$ and $\mathbf{m}_i$ are the covariance matrix and mean vector of the pattern feature of class ω_i, and $\det \mathbf{C}_i$ is the determinant of $\mathbf{C}_i$. Expression (13) is used to determine the fuzzy density for each test sample that will be discussed in the next section

Using the monotonically increasing property of the logarithm, the decision function $d_i(\mathbf{x})$ has the following logarithmic form

$$d_i(\mathbf{x}) = \ln[p(\mathbf{x}|\omega_i)P(\omega_i)] = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i) \qquad (14)$$

The substitution of the expression for the Gaussian probability distribution function expressed in (13) into (14) and after some mathematical rearrangement give

$$\begin{aligned} d_i(\mathbf{x}) \quad = \quad & \ln P(\omega_i) - \frac{1}{2}\ln(\det \mathbf{C}_i) \\ & -\frac{1}{2}[(\mathbf{x}-\mathbf{m}_i)^T \mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i)] \qquad (15) \end{aligned}$$

The equation expressed in (15) is known as the Bayesian decision function for Gaussian pattern class ω_i under the condition of a 0-1 loss function.

5 FUSION-BASED CLASSIFICATION OF CELLULAR MORPHOLOGIES USING PHENOTYPE FEATURES

Fluorescent microscopy images of cells stained to reveal complex cellular features, such as cytoarchitecture, are considered to be high-content images due to the large amount of information they contain. These images reveal numerous biological readouts, including cell size, cell viability, DNA content, cell cycle, and cell morphology. A gene's function can be assessed by analyzing alterations in a biological process caused by the absence of that gene. A specific study concerns a cell-based assay for the activity of the *Rho GTPase Rac1* using the Drosophila *Kc167* embryonic cell line (Wang et al., 2008).

Distinct morphological changes in cells both in vitro and in vivo caused by constitutively active forms of *Rho* proteins can be observed. *Kc167* cells are small and uniformly round.

We used 643 *normal*, 321 *ruffling*, and 210 *spiky* RNAi cell samples in this study. Using a feature extraction procedure (Wang *et al.*, 2008), 211 texture and shape features for each cell image were obtained.

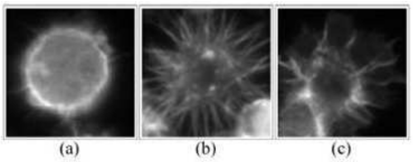

Figure 1: Three cellular phenotypes of Drosophila Kc167 cells: a) Normal; b) Spiky; c) Ruffling.

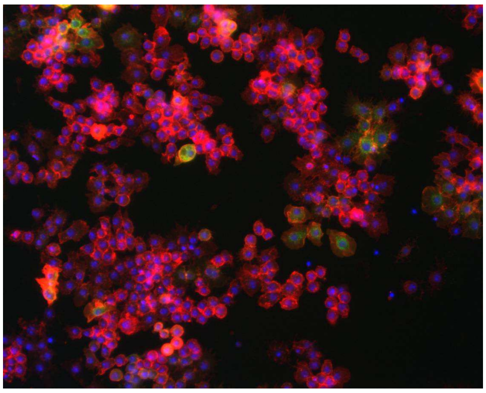

Figure 2: Typical screening image.

To test the performance of the proposed fuzzy information fusion, we used only 2 sets of features, each set consists of 12 different features of the 211 textures and shapes previously discussed. We used half of the samples for data testing (320, 100, and 160 for class 1, class 2, and class 3; respectively) and the other samples were used for training. Let A be a target class, and B and C the two different feature sets used to identify A. For each set of features, expression (13) was used to compute $p(\mathbf{x}|\omega_i)$ for each sample $\mathbf{x}$ for class ω_i, giving values for $P(A|B) = g(A|B)$ and $P(A|C) = g(A|C)$ for each of the 3 classes (*normal*, *spiky*, and *ruffling*). Prior probabilities or fuzzy densities for A for each of the 3 classes are assumed to be equal, giving $P(A) = g(A) = 1/3$. We consider that these sets of feature vectors do not completely cover all possible features associated with the phenotypes. Therefore, the subset ϕ was introduced to represent the set of all remaining possible features. The fuzzy density of ϕ can be subjectively estimated. In this study we set $g(\{\phi\}) = 0.4$ to represent the degree of belief of the phenotype existence when all other unforseen features are considered. It is noted that using different reasonable values of $g(\{\phi\})$ will not affect the relative comparisons of the interactions of the actual features. The Bayes classifier was used to classify the testing samples based on each of the two feature sets. The two sets of output obtained from the Bayes classifier were then combined by the ratio-based (probabilistic) fusion and the ratio-based fuzzy fusion operators.

Table 1: Correction rates (%) on classification of cellular phenotypes using different methods.

Class	1	2	3
Bayes classifier 1	66.25	62.00	63.12
Bayes classifier 2	64.06	64.00	61.25
Ratio-based fusion	76.88	65.00	65.00
Ratio-based fuzzy fusion	80.31	74.00	66.25

Table 2: Confusion matrix of Bayes classifier 1.

Class	1	2	3	Total
1	212	39	69	320
2	18	62	20	100
3	27	32	101	160

Table 3: Confusion matrix of Bayes classifier 2.

Class	1	2	3	Total
1	205	44	71	320
2	14	64	22	100
3	26	36	98	160

Table 4: Confusion matrix of probabilistic fusion.

Class	1	2	3	Total
1	246	25	49	320
2	16	65	19	100
3	24	32	104	160

Table 5: Confusion matrix of fuzzy fusion.

Class	1	2	3	Total
1	257	24	39	320
2	11	74	15	100
3	24	30	106	160

The classification rates obtained from Bayes classifier 1 (using feature set 1), Bayes classifier 2 (using feature set 2), probabilistic fusion, and fuzzy fusion are given in Table 1. The confusion matrices (non-diagonal elements indicate misclassified samples) of the Bayes classifier 1, Bayes classifier 2, probabilistic fusion, and fuzzy fusion are given in Tables 2-5, respectively. The experimental results show that the combined results are better than those obtained from individual classifiers, and suggest the best performance of the proposed fuzzy approach in all classes.

6 CONCLUSIONS

A proposed fusion scheme and its preliminary application for classifying phenotypic classes of biological cells have been discussed. The initial result for validating the proof of concept of the model seems to be promising for combining results from various sources.

Extended investigation of the proposed approach is under way to develop a key component for automatic cellular phenotype identification as an effort toward the construction of a robust automated imaging system for high-content genome-wide screening.

REFERENCES

Banon G. Distinction between several subsets of fuzzy measures, Fuzzy Sets and Systems 1981, 5: 291-305.

Berner ES, Ball MJ, Hannah KJ, Eds. Clinical Decision Support Systems: Theory and Practice. New York: Springer-Verlag, 1998.

Chi Z, Yan H, Pham, T. Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition. Singapore: World Scientific, 1996.

Das S. High-Level Data Fusion. Norwood: Artech House, 2008.

Gonzalez RC, Woods RE. Digital Image Processing. NJ: Prentice Hall, 2002.

Grabisch M. The representation of importance and interaction of features by fuzzy measures, Pattern Recognition Letters 1996, 17: 567-575.

Journel AG. Combining knowledge from diverse sources: An alternative to traditional data independence hypotheses, *Mathematical Geology*, 34 (2002) 573-595.

Klir GJ, Wierman MJ. Uncertainty-Based Information: Elements of Generalized Information Theory. Heidelberg: Physica-Verlag, 1999.

Lesczynski K, Penczek P, Grochulski PW. Sugeno's fuzzy measure and fuzzy clustering, Fuzzy Sets and Systems 1985, 15: 147-158.

Muller C, Rombaut M, Janier M. Dempster Shafer approach for high level data fusion applied to the assessment of myocardial viability, T. Katila et al. (Eds.): LNCS 2230, pp. 104-112, 2001.

Pellizzeri TM, Lombardo P, Oliver CJ. A new maximum likelihood classification technique for multitemporal SAR and multiband optical images, Proc. IGARSS 2002, pp. 24-28

Shafer GA. A Mathematical Theory of Evidence. NJ: Princeton University Press, 1976.

Shortliffe EH. Computer-based Medical Consultations: Mycine. New York: Elsevier, 1976.

Sugeno M. Fuzzy measures and fuzzy integrals: A survey, in: Fuzzy Automata and Decision Processes, Amsterdam: Elsevier, pp. 89-102, 1977.

Tahani H, J. Keller J. Information fusion in computer vision using the fuzzy integral, IEEE Trans. Systems, Man, and Cybernetics 1990, 20: 733-741.

Wang J, Zhou X, Bradley PL, Chang SF, Perrimon N, Wong STC. Cellular phenotype recognition for high-content RNA interference genome-wide screening, J. Biomolecular Screening 2008, 13: 29-39.