Information Fusion 18 (2014) 9–19

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Quantum inspired method of feature fusion based on von Neumann entropy

Weimin Peng [a,b], Huifang Deng [a,*]

[a] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China
[b] School of Information, Guangdong Ocean University, Zhanjiang 524088, Guangdong, China

ARTICLE INFO

Article history:
Received 24 July 2012
Received in revised form 12 October 2013
Accepted 12 October 2013
Available online 26 October 2013

Keywords:
Data fusion
Feature fusion
Duplicate detection
Quantum information
Von Neumann entropy

ABSTRACT

One mission of feature fusion is to obtain a complete yet concise presentation of all existing feature data by detecting and fusing the duplicate feature data. In contrast to the already developed feature fusion methods which have shown their limitations, this paper applies the theories of quantum information to feature fusion. Further, a novel and effective step-wise quantum inspired feature fusion method, which detects the duplicate feature data based on maximum von Neumann mutual information and fuses the duplicate feature data using the operations on quantum state, is developed. This same idea is also used for feature dimensionality reduction, and the corresponding models are investigated. For comparison, another quantum inspired feature fusion method based on average quantum phase is presented here. The experimental results show that the quantum inspired feature fusion method based on von Neumann entropy gives better results on completeness and conciseness than the method based on average quantum phase.

1. Introduction

With widespread availability of the Internet and the rapid development of the Internet of Things [1] which interlinks any physical objects in the real world, heterogeneous mass data need processing. While many data records actually represent the same real world object among the mass data. For example, fifteen students are invited to write a same word ("brown") which is inputted into a digital dataset. Obviously, these fifteen words represent the same real word object, so, a duplicate record detection is needed. Otherwise, data redundancy becomes inevitable, thus, the conciseness which we are pursuing is deteriorated and the contradiction in data processing may occur. So, one mission of data fusion is to fuse the multiple records into a single, consistent and clean representation [2] as shown in Fig. 1, which is a key step of data integration. Sorted by the data source, data fusion is classified into decision fusion, feature fusion and signal fusion. Signal fusion operates on the raw data and it is a low level fusion. Feature fusion operates on the feature data which is extracted from the raw data and it is a medium level fusion. While decision fusion operates on the preliminary decision data which have gone through the processes of feature extraction, identification, and decision on the basis of the raw data. In this paper, we concentrate on feature fusion. Our main purpose is to improve the completeness and conciseness of fusion results.

Before feature fusion, one needs to do feature schema mapping and duplicate feature detection [3]. For example, in order to digitalize the handwritten words in Fig. 1, digital schemas are needed. In general, the schemas to digitalize the source objects are called as source schemas and the target schema is used to digitalize the fusion result. Normally, the target schema and the source schemas are different, so, mapping the source schemas to the target schema is needed. As the source schemas in Fig. 1 represent the same target object, therefore, detection of the duplicate source objects after digitalization is also needed. In order to concentrate on feature fusion, we assume that the source feature data possess the same target feature schema. Thus, our main task is to detect the duplicate data and to fuse them into a single representation. Although the target feature schema is the same, different observing techniques may lead to inconsistent or contradictive feature values [4] representing the same feature object. In this paper, we shall ignore this issue.

Usually one uses effectiveness and efficiency [2] to measure the performance of duplicate feature detection. Generally, the quality of the detection principle and threshold has great impacts on effectiveness. In a digitalization society, to store and calculate large datasets is a formidable task, so, efficiency in duplicate detection is an issue one has to address. The performance of feature data fusion includes completeness, the number of unique objects in a dataset in relation to the overall number of unique objects in the real world, and conciseness, the number of unique objects in a dataset in relation to the overall number of objects in the dataset [2].

* Corresponding author.
E-mail addresses: pengwm@gmail.com (W. Peng), hdeng2008@gmail.com (H. Deng).

1566-2535/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.inffus.2013.10.003

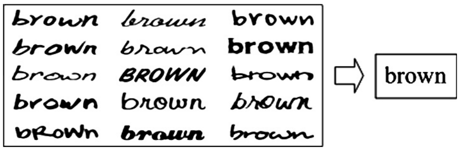

Fig. 1. Fusing fifteen handwritten words into a single representation.

There are two main types of macro-techniques of feature fusion: joint-based technique and union-based technique. Joint-based technique fuses the feature records from different sources into a super-record. It performs well in terms of conciseness [2]. Union-based technique fuses all the records from different sources into the results without any loss of source information. It performs well in terms of completeness [2]. In this paper, we use joint-based technique as macro-fusion technique.

Ref. [5] presented the known methods, algorithms, architectures, and models of information fusion, and discussed their applicability in the context of wireless sensor networks. Sometimes, data fusion is also termed as information fusion. The known specific techniques used for data fusion are divided into two categories: inference-based techniques, such as Dempster–Shafer inference [6], Bayesian inference [7,8], fuzzy logic [9], and neural networks [10–12], and estimation-based techniques, such as maximum likelihood (ML) [13], least squares [14], Kalman filter [15], and particle filter [16]. The known feature fusion methods include feature concatenation [17], combinational classifiers [18] and kernel-based methods, such as semi-definite programming [19], Bayesian hierarchical models [20–22]. In Ref. [23], LLE [24] technique is used to fuse feature data for classification and dimensionality reduction. Quantum computation and quantum information deal with the information processing that can be accomplished using quantum mechanics, which is a mathematical framework or set of rules for the construction of physical theories [25]. In this paper we apply some ideas of quantum information theory to feature fusion and attempt to develop a novel quantum inspired feature fusion method.

In 1994, Shor [26,27] proposed a quantum factorization algorithm with polynomial time complexity. This algorithm declares the era of quantum computing. Ref. [28] introduced quantum computing to non-physicists and established the theoretical and technical basis of quantum computing. The book, Quantum Computation and Quantum Information [25], is widely regarded as a bible of quantum computing and quantum information. Ref. [29] also elaborated the details of quantum computing and quantum information. With regard to dense quantum coding, Ref. [30] investigated the possibility of encoding m classical bits into many fewer n quantum bits so that an arbitrary bit from the original m bits can be recovered with good probability. Ref. [31] investigated optimal encoding and retrieval of digital data, when the storage/communication medium is described by quantum mechanics. While regards von Neumann entropy, Ref. [32] proposed a novel method for integrating different biological data sets based on the entropy of kernel matrices. According to the calculation methods of quantum information, in this paper, we study feature fusion based on von Neumann entropy and von Neumann mutual information [29], and make a comparison with another quantum inspired method based on average quantum phase (AQP), which is designed by borrowing the idea of discretization [33]. Our main purpose is to improve the effectiveness of duplicate detection and the completeness and conciseness of fusion results.

In Section 2, the mathematical models of feature fusion based on von Neumann entropy are presented, including the mathematical models of duplicate detection. In Section 3 the mathematical model of feature dimensionality reduction is formulated according to the idea of the feature fusion method based on von Neumann entropy. The relevant experimental results are shown in Section 4. The conclusions are drawn in the last Section.

2. Mathematical models of feature fusion

A dataset is the collection of the real feature vector $X^i = (x_1^i, x_2^i, \ldots, x_j^i, \ldots, x_n^i)^T$, which includes n feature samples. Where, $1 \leqslant i \leqslant L$ and L is the number of features. In general, there exists one class feature which classifies the dataset into several sub-datasets. Assuming that the Lth feature is the class feature, we first divide the sample space into different subspaces according to the value of class feature. The models for subsequent duplicate detection and feature fusion are the same in each subspace, so we just consider the models in one subspace. In this sense, n refers to the number of samples in that subspace.

Assuming that the interval of the real element x_j^i in X^i is $[a, b]$ and the value increment of x_j^i equals 1, then the corresponding basic quantum state $|q_j^i\rangle$ is represented as

$$|q_j^i\rangle = |Q_{j'}^i\rangle = |b_1\ b_2 \cdots b_k \cdots b_{m_i}\rangle \tag{1}$$

where $|Q_{j'}^i\rangle$ is one of the unique quantum states which constitute $|X^i\rangle$, $0 \leqslant j' \leqslant (b-a)$ and j' is equal to $x_j^i - a$, $|b_k\rangle$ is equal to $|0\rangle$ or $|1\rangle$, and m_i is the length of qubits which is equal to

$$m_i = \lceil \log_2(b-a+1) \rceil \tag{2}$$

So the number of unique quantum states is 2^{m_i} and the probability amplitudes of the unique quantum states $|Q_1^i\rangle, |Q_2^i\rangle, \ldots, |Q_{j'}^i\rangle, \ldots, |Q_{2^{m_i}}^i\rangle$ are denoted as $\mu_1^i, \mu_2^i, \ldots, \mu_{j'}^i, \ldots, \mu_{2^{m_i}}^i$. They must satisfy the requirement of normalization, that is, $|\mu_1^i|^2 + |\mu_2^i|^2 + \cdots + |\mu_{j'}^i|^2 + \cdots + |\mu_{2^{m_i}}^i|^2 = 1$. In other words, the quantified feature vector $|X^i\rangle$ is the superstition quantum state of all the unique quantum states and the probability of collapsing to the quantum state $|Q_{j'}^i\rangle$ is $|\mu_{j'}^i|^2$ when $|X^i\rangle$ is measured. As the interval of x_j^i is $[a, b]$, it is easy to show that the number of probability amplitudes with value of zero is at least $2^{m_i} - (b-a+1)$. As a result, we expresses $|X^i\rangle$ as a quantum superposition state.

$$|X^i\rangle = \mu_1^i|Q_1^i\rangle + \mu_2^i|Q_2^i\rangle + \cdots + \mu_{j'}^i|Q_{j'}^i\rangle + \cdots + \mu_{2^{m_i}}^i|Q_{2^{m_i}}^i\rangle \tag{3}$$

For example, if the interval of the real elements in the vector X^1 is [0, 100], then the length of qubits m_1 equals 7, and the number of basic quantum states is 2^7, i.e., 128. Since the number of real elements in X^1 is 101, there are at least 27 basic quantum states having probability amplitudes of zero. Assuming that x_6^1 is a real element in X^1 whose value equals 32, then, the corresponding basic quantum state $|q_6^1\rangle$ is denoted as $|q_6^1\rangle = |Q_{32}^1\rangle = |0100000\rangle$. As we know, 0100000 is the binary expression of 32. Obviously, the probability of $|q_6^1\rangle$ is not equal zero and it is determined by its occurrence frequencies (i.e., times) in $|X^1\rangle$.

Based on the quantum representation, we count the frequency of each unique quantum state in the vector $|X^i\rangle$. If the known unique quantum sate $|Q_{j'}^i\rangle$ appears $o_{j'}^i$ times in $|X^i\rangle$ then the measurement probability of $|Q_{j'}^i\rangle$ equals

$$p_{j'}^i = (\mu_{j'}^i)^2 = o_{j'}^i / (o_1^i + o_2^i + \cdots + o_{j'}^i + \cdots + o_{2^{m_i}}^i) \tag{4}$$

In quantum mechanics, the density matrix carries equivalent information to the quantum state. The density matrix $\rho_{j'}^i$ of the known quantum state $|Q_{j'}^i\rangle$ is defined as

$$\rho_{j'}^i = p_{j'}^i |Q_{j'}^i\rangle\langle Q_{j'}^i| \tag{5}$$

So, the density matrix of the ensemble of all the known quantum states in the feature vector $|X^i\rangle$ is denoted as

$$\rho^i = \sum_{j'} p_{j'}^i |Q_{j'}^i\rangle\langle Q_{j'}^i| \tag{6}$$