Information Fusion 31 (2016) 127–136

Contents lists available at ScienceDirect

Information Fusion

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

An intelligent quality-based approach to fusing multi-source probabilistic information

Ronald R. Yager [a,*], Fred Petry [b]

[a] Machine Intelligence Institute, Iona College, New Rochelle, NY 10801, United States
[b] Marine Geosciences Division, Geospatial Sciences & Technology Branch, Naval Research Laboratory, Stennis Space Center, MS 39529, United States

ARTICLE INFO

Article history:
Received 9 July 2015
Revised 3 February 2016
Accepted 7 February 2016
Available online 16 February 2016

Keywords:
Fusion
Entropy
Credibility
Quality-based

ABSTRACT

Our objective here is to obtain quality-fused values from multiple sources of probabilistic distributions, where quality is related to the lack of uncertainty in the fused value and the use of credible sources. We first introduce a vector representation for a probability distribution. With the aid of the Gini formulation of entropy, we show how the norm of the vector provides a measure of the certainty, i.e., information, associated with a probability distribution. We look at two special cases of fusion for source inputs those that are maximally uncertain and certain. We provide a measure of credibility associated with subsets of sources. We look at the issue of finding the highest quality fused value from the weighted aggregations of source provided probability distributions.

1. Introduction

The use of fusion to combine data provided by multiple sources about the value of a variable is common in many applications [1]. One rational for fusing probabilistic distributions provided by multiple sources is to improve the quality of the information to decision makers [2]. Our interest here is looking at the problem of obtaining high quality fused values. One aspect of this quality is a reduction in the uncertainty of the information. Unfortunately, combining probability distributions information does not always result in a probability distribution with less uncertainty, this particularly is the case when the data that are being fused is conflicting. In order to formally quantify the uncertainty associated with a probability distribution we will use the concept of entropy. A second contributing factor to the association of quality with a fused value is that we have used quality sources of information, the more of these sources used, the more credible the results of the fusion process. In order to capture this criterion of a quality fusion we introduce a measure of credibility associated with use of various subsets of the sources. Here we provide a quantification of the notion of a quality fusion based on the objective of providing fused values having little uncertainty based on a credible subset of the sources.

2. Vector representation of probability distributions

Assume P_i is a probability distribution on the space $X=\{x_1,\ldots, x_n\}$ where p_{ij} is the probability of the occurrence of x_j. Here, each $p_{ij} \in [0, 1]$ and $\sum_{j=1}^{n} p_{ij} = 1$. For our purposes in the following we shall find it useful, at times, to represent a probability distribution as an n-dimensional vector $P_i=[p_{i1}, p_{i2},\ldots, p_{in}]$. Here the vector has the special properties that all its components lie in the unit interval and their sum is one.

If P_i for $i=1$ to q are a collection of probability distribution vectors then their weighed sum, $P = \sum_{i=1}^{q} w_i P_i$, is another vector whose components are $p_j = \sum_{i=1}^{q} w_i p_{ij}$. Furthermore, if the weights are standard weights, $w_i \in [0, 1]$ and $\sum_{i=1}^{q} w_i = 1$, then P is also a probability distribution vector.

Another operation on vectors is the dot or inner product, see Bustince and Burillo [3]. If P_i and P_k are two probability vectors on the space X then their dot product is

$$P_i \cdot P_k = \sum_{j=1}^{n} p_{ij} p_{kj}.$$

We emphasize that the dot product is a scalar value. Furthermore, in the case where P_i and P_k are probability distributions then $0 \leq P_i \cdot P_k \leq 1$. A special case of dot product is where P_i and P_k are the same then $P_i \cdot P_k = \sum_{j=1}^{n} (p_{ij})^2$. For notational simplicity at times when it causes no confusion, we shall simply use $P_i P_k$ for the dot product.

An important concept that is associated with this self dot product is the idea of the norm of the vector. In particular then norm

* Corresponding author. Tel.: +1 212 249 2047.
E-mail addresses: yager@panix.com (R.R. Yager), fred.petry@nrlssc.navy.mil (F. Petry).

http://dx.doi.org/10.1016/j.inffus.2016.02.005
1566-2535/

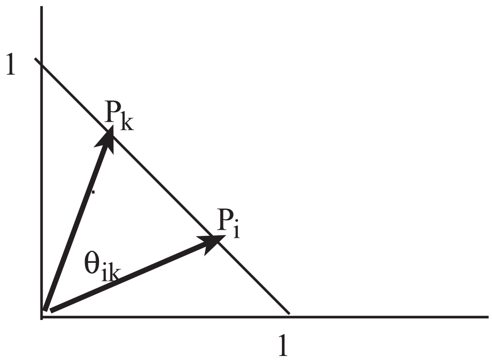

Fig. 1. Angle between probabilistic vectors.

$\|P_i\| = \sqrt{P_i P_i} = (\sum_{j=1}^{n} (p_{ij})^2)^{1/2}$. The norm is referred to as the Euclidean length of a vector. Because of the special properties of the probability distribution vector, $p_{ij} \in [0, 1]$ and $\Sigma_i p_{ij} = 1$, it can be easily shown that the maximal value of $\|P_i\|$ occurs when one $p_{ij} = 1$ and all other $p_{ij} = 0$. In this case, $\|P_i\| = 1$. Furthermore, in this case of a probability distribution vector the minimum value of $\|P_i\|$ occurs when all $p_{ij} = 1/n$ and this has the value $\|P_i\| = (\sum_{i=1}^{n} (\frac{1}{n})^2)^{1/2} = (\frac{1}{n})^{1/2} = \frac{1}{\sqrt{n}}$. We note for the self dot product, $P_i P_i = \|P_i\|^2$ we have a maximal value of one and minimal value of $\frac{1}{n}$ when all $p_{ij} = \frac{1}{n}$.

In the following we shall benefit from the use of an illustration of the probability vector in the two-dimensional case as shown in Fig. 1.

If P_i and P_K are two probability vectors it is known [4] that the Cosine of the angle between them denoted θ_{ik} is expressed as

$$\cos(\theta_{ik}) = \frac{P_i P_k}{||P_i||||P_k||}$$

We note $\cos(\theta_{ik})$ is the dot product of P_i and P_k divided by their respective norms. It is well known that if $\cos(\theta_{ik}) \in [0, 1]$, as is the case when P_i and P_k are probability distribution vectors, that $\theta_{ik} \in [0, \frac{\pi}{2}]$.

We further see that if $P_i = P_k$ then $\cos(\theta_{ik}) = \frac{P_i P_k}{||P_i|| \, ||P_k||} = \frac{P_i^2}{||P_i||^2} = \frac{P_i^2}{P_i^2} = 1$. Thus if P_i and P_k are the same, coincident, then $\cos(\theta_{ik}) = 1$. Furthermore it is known in this case that $\theta_{ik} = 0$. At the other extreme is the case where P_i and P_k are orthogonal, $P_i P_k = \sum_{j=1}^{n} p_{ij} p_{kj} = 0$ where $\cos(\theta_{ik}) = \frac{P_i P_k}{||P_i|| \, ||P_k||} = 0$. We get in this case that $\theta_{ik} = \frac{\pi}{2}$. We note that in the case where P_i and P_k are orthogonal then $p_{ij} = 0$ when $p_{ik} \neq 0$ and $p_{ik} = 0$ when $p_{ij} \neq 0$.

We illustrate these extremes of coincident and orthogonal distributions for the two dimensional case in Fig. 2.

We note in the n-dimensional case a prototype example of orthogonality occurs when P_i has $p_{ij1} = 1$ and P_k is $p_{kj2} = 1$. Here they each completely support different outcomes.

In [5] we suggested that $\cos(\theta_{ik})$ can be used as measure of the degree of compatibility, Comp, between the two probability distributions, thus

$$\text{Comp}(P_i, P_k) = \frac{P_i P_k}{||P_i|| \, ||P_k||}$$

Here $\text{Comp}(P_i, P_k) \in [0, 1]$ and the closer to one the more compatible the probability distributions. Furthermore $1 - \text{Comp}(P_i, P_k)$, denoted $\text{Conf}(P_i, P_k)$, can be seen as the degree of conflict between the two probability distributions. We note that if P_i and P_k are orthogonal then $\text{Comp}(P_i, P_k) = 0$ that $\text{Conf}(P_i, P_k) = 1$. On the other hand if P_i and P_k are coincident, the same, then $\text{Comp}(P_i, P_k) = 1$ and $\text{Conf}(P_i, P_k) = 0$.

An interesting special case occurs when one of the distributions, P_i, has $p_{ij} = \frac{1}{n}$ for all j. Here we previously noted $\|P_i\| = (\frac{1}{n})^{1/2}$. Consider now $\text{Comp}(P_i, P_k)$ where P_i is this uniform probability distribution. Here $\text{Comp}(P_i, P_k) = \frac{P_i P_k}{||P_i|| \, ||P_k||}$. However in this case

$$P_i P_k = \sum_{j=1}^{n} p_{ij} p_{kj} = \frac{1}{n} \sum_{i=1}^{n} p_{kj} = \frac{1}{n}$$

and thus $\text{Comp}(P_i, P_k) = \frac{\frac{1}{n}}{||P_k||(\frac{1}{n})^{1/2}} = \frac{(\frac{1}{n})^{1/2}}{||P_k||} = \frac{1}{\sqrt{n}} = \frac{1}{||P_k||}$. Two special cases of P_k are worth commenting on. If P_k is a certain distribution, it has $p_{kj} = 1$ for one element, then $||P_k|| = 1$ and $\text{Comp}(P_i, P_k) = \frac{1}{\sqrt{n}}$. If P_k is also a uniform probability distribution, all $p_{kj} = \frac{1}{n}$, here then $||P_k|| = \frac{1}{\sqrt{n}}$ and we get $\text{Comp}(P_i, P_k) = 1$.

3. Entropy, certainty and information

An important concept associated with a probability distribution on the space $X = \{x_1, \ldots, x_n\}$ is the idea of entropy [6,7]. The most common measure of entropy is the Shannon entropy. Here if P is a probability distribution on the space with p_j the probability associated with x_j, then the Shannon entropy is $H(P) = -\sum_{j=1}^{n} p_j \ln(p_j)$. It is well known that the maximal entropy occurs when all $p_i = \frac{1}{n}$ in which case $H(P) = \ln(n)$. The minimal entropy occurs for the case when one $p_j = 1$ and all other $p_j = 0$, in this case $H(P) = 0$. What is clear is that the entropy is measuring the uncertainty associated with the probability distribution, the more uncertainty the more entropy. The complimentary idea of entropy is certainty (or

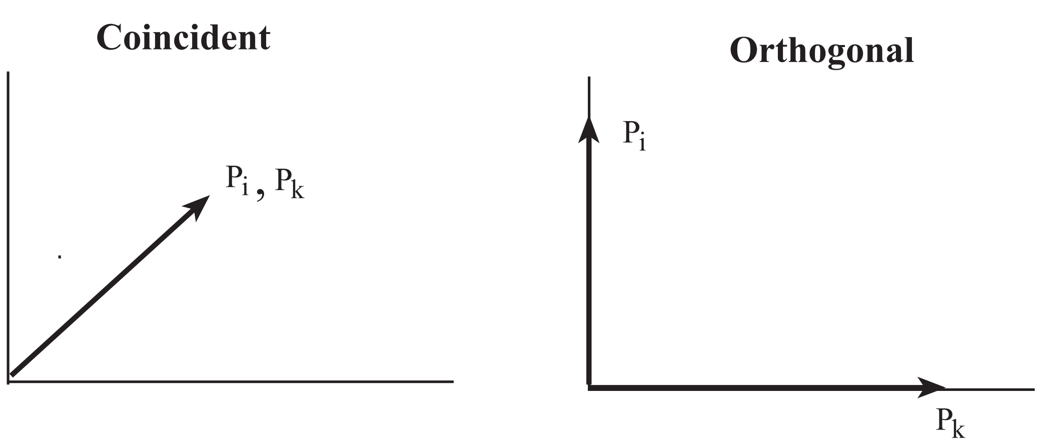

Fig. 2. Different relationships between probabilistic distributions.