VARIABLE PROBABILITY-POSSIBILITY TRANSFORMATION

Moamar Sayed Mouchaweh, Mohamed Saïd Bouguelid, Patrice Billaudel, Bernard RIERA

Université de Reims, CReSTIC - LAM, Moulin de la Housse B.P. 1039, 51687 REIMS Cedex 2, France
moamar.sayed-mouchaweh@univ-reims.fr, 33 (0) 3 26 91 82 17 - 33 (0) 3 26 91 31 06

Abstract

Our team of research works on diagnosis by pattern recognition using Fuzzy Pattern Matching (FPM) as a classification method, for data coming from industrial and medical sectors. FPM uses a transformation from probability to possibility in order to construct densities of possibilities. These densities are used to assign each new sample to its suitable class which corresponds to a functioning mode. In this paper, a transformation from probability to possibility is proposed. This transformation is named Variable Transformation because its specificity property varies with a parameter. The performances of the Variable Transformation are compared with the ones of some well known ones of the literature using some meaningful criteria. We show that the Variable Transformation has interesting characteristics for the diagnosis by pattern recognition, in particular that it can be the most informative transformation, in the discrete case, and it is the one that best distinguishes the confused elements.

Keywords: Fuzzy logic, Diagnosis by pattern recognition, Possibility theory, Probability-possibility transformations

1 Introduction

Artificial Intelligence systems are constrained to deal with imperfect data, and thus to reason approximately under conditions of ignorance. We consider three main theories namely: Probability, Possibility and Evidence theories, as alternative frameworks to represent and to handle imperfect data. Probability theory is built on five axioms that define the behaviour of a probability measure, which may be used as an estimate of degree to which an uncertain event is likely to occur (Parsons 1998). The Probability theory is inadequate for dealing with imprecise knowledge such as “Mark is young” and with natural language expression of uncertainty such as “likely” (Walley 1996). The Possibility theory, which was founded by (Zadeh 1978), is a natural tool to model the imprecision and the uncertainty when they occur together in ordinary language such as “It is likely that Mark is young”. The Evidence theory, which was initiated by Dempster and developed by (Shafer 1976), includes both the Probability and the Possibility theories. It uses two related measures, called belief and plausibility measures, as upper and lower probabilities. The probability and possibility measures are special cases of the belief and plausibility measures (Oussalah 2000).

There exist different types of data so; depending on available data it is necessary to choose the appropriate theory. The problem arises when we have different types of data represented in different theories and we have to handle or combine them. In this case a transformation from one theory to another is useful. Transformation from probability into possibility is useful for some practical problems as: constructing a fuzzy membership function from statistical data (Krishnapuram 1993), (Mouchaweh 2004), (Devillez 2004) combining probabilities and possibilities information in expert systems (Klir 1992) and reducing the computational complexity (Dubois 1993). In other hand, the transformation from possibility into probability is useful in the case of decision making when the expert needs precise information to take his decision.

There are different probability-possibility transformations in the literature. Their performances can be evaluated using some meaningful criteria. In this paper, we introduce the main probability-possibility consistency conditions and investigate the most known transformations in the literature. We propose a transformation from probability to possibility, which we name Variable Transformation (VT), and compare its performances with the ones of the previous transformations. We show that VT has interesting characteristics for the diagnosis by pattern recognition, in particular that it can be the most informative transformation, in the discrete case, and it is the one that best distinguishes the confused elements.

2 Transformation consistency principles

2.1 Zadeh consistency principle

Zadeh defined the probability-possibility consistency principle such as "a high degree of possibility does not imply a high degree of probability, nor does a low degree of probability imply a low degree of possibility" (Zadeh 1978). He defined the degree of consistency between a probability distribution $p = (p_1, p_2, \ldots, p_n)$ and a possibility distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ as:

$$C_Z = \sum_{i=1}^{n} \pi_i . p_i \qquad (1)$$

Zadeh pointed out that the probability-possibility consistency, defined in (1), is not a precise law or a relationship between possibility and probability distributions. It is an approximate formalization of the heuristic connection that a lessening of the possibility of an event tends to lessen its probability but not vice-versa (Zadeh 1978).

2.2 Klir consistency principle

Let $X = \{w_1, w_2, \ldots, w_n\}$ be a finite universe of singletons, let $p_i = p_i(w_i)$ and $\pi_i = \pi(w_i)$. Assume that the elements of X are ordered in such a way that: $\forall\ i = 1 \,..\, n : p_i > 0$, $p_i \geq p_{i+1}$ and $\pi_i > 0$, $\pi_i \geq \pi_{i+1}$ with $p_{n+1} = 0$ and $\pi_{n+1} = 0$. We will consider this notation for all the formulas of this paper. According to Klir, the transformation from p_i to π_i must preserves some appropriate scale and the amount of information contained in each distribution (Klir 1993). The information contained in p or π can be expressed by the equality of their uncertainties. Klir has considered the principle of uncertainty preservation under two scales:

– The ratio scale: which is a normalization of the probability distribution? The transformations $p \rightarrow \pi$ and $\pi \rightarrow p$ are named the normalized transformations and they are defined by:

$$\pi_i = \frac{p_i}{p_1}\,,\ p_i = \frac{\pi_i}{n.\sum_{i=1}^{n} \pi_i} \qquad (2)$$

– The log-interval scales: the corresponding transformations $p \rightarrow \pi$ and $\pi \rightarrow p$ are defined by:

$$\pi_i = \left(\frac{p_i}{p_1}\right)^{\alpha},\ p_i = \frac{\pi_i^{1/\alpha}}{\sum_{i=1}^{n} \pi_i^{1/\alpha}} \qquad (3)$$

These transformations, which are named Klir transformations, satisfy the uncertainty preservation principle defined by (Klir 1993). α is a parameter that belongs to the open interval $]0, 1[$.

2.3 Dubois and Prade consistency principle

The possibilistic representation is weaker than the probabilistic one because it explicitly handles imprecision (e.g. incomplete data) and because possibility measures are based on ordering structure than an additive one in the probability measures (Dubois 1993). Thus in going from a probabilistic representation to a possiblistic one, some information is lost because we go from point-valued probabilities to interval valued ones; the converse transformation adds information to some possibilistic incomplete knowledge.

The transformation $p \rightarrow \pi$ is guided by the principle of maximum specificity, which aims at finding the most informative possibility distribution. While the transformation $\pi \rightarrow p$ is guided by the principle of insufficient reason which aims at finding the probability distribution that contains as much uncertainty as possible but that retains the features of possibility distribution (Dubois 1993). This leads to write the consistency principle of Dubois and Prade such as:

$$\forall A \subset X : \Pi(A) \geq P(A) \qquad (4)$$

The transformations $p \rightarrow \pi$ and $\pi \rightarrow p$ are defined by:

$$\pi_i = \sum_{j=i}^{n} p_j \ , \ p_i = \sum_{j=i}^{n} \frac{(\pi_j - \pi_{j+1})}{j} \qquad (5)$$

The two transformations defined by (4) and (5) are not converse of each other because they are not based on the same informational principle. For this reason, we name the transformation $p \rightarrow \pi$ defined by (5) as asymmetric one. Dubois and Prade suggested a symmetric $p \rightarrow \pi$ transformation which is defined by:

$$\pi_i = \sum_{j=i}^{n} min(\ p_i, p_j\ ) \qquad (6)$$

Dubois and Prade proved that the asymmetric transformation $p \rightarrow \pi$, defined by (6), is the most specific transformation which satisfies the condition of consistency of Dubois and Prade defined by (4) (Dubois 1993).

2.4 Variable transformation

We propose the following parametric transformation $p \rightarrow \pi$ which we name the Variable Transformation (VT):

$$\pi_i = \left(\frac{p_i}{p_1}\right)^{k.(1-p_i)} \qquad (7)$$

k is a constant which must guarantee the following condition of consistency:

$$\forall\ w \in X : \pi(w) \geq p(w) \qquad (8)$$

This condition, which is a particular case of Dubois and Prade consistency principle defined by (4), entails a possibility distribution with interesting characteristics for the diagnosis by pattern recognition as we will see in the next section. Indeed (8) is the discrete case, i.e. distribution containing a set of singletons, of the consistency condition of Dubois and Prade defined by (4). To guarantee the consistency principle defined by (8), the value of k must belong to the following interval (Mouchaweh 2002):

$$0 \leq k \leq \frac{\log p_n}{(1-p_n).\log\frac{p_n}{p_1}} \qquad (9)$$

When the value of k is equal to its maximum value: $k_{max} = \frac{\log p_n}{(1-p_n).\log\frac{p_n}{p_1}}$, the possibility π_n, calculating according to the VT, is equal to p_n . If we increase the value of k above k_{max} then the value of π_n becomes less than the one of p_n which means that the VT does not satisfy any more the consistency principle defined by (8) for all $i = 1,\ldots,n$. Indeed, the VT transforms a probability distribution in a non-linear way; this means that the VT adds more of information to the high probabilities than to the small ones. This is due to the fact that high probabilities have a power, $k.(1-p_i)$, smaller than the one of small probabilities. The difference between Klir transformation and the VT is that Klir

transformation has a constant power α which belongs to the open interval]0, 1[in order to preserve the uncertainty, while the power, $k.(1-p_i)$, in the VT, is variable to make it more specific.
In the next section, we compare the VT performances with the ones of the previous transformations in using some evaluation criterions. We prove that the VT has interesting characteristics for diagnosis by pattern recognition problems.

3 Evaluation criterions

3.1 Ignorance preservation criterion

This criterion aims to preserve the complete ignorance state. That is, if the initial distribution is close to the complete ignorance case, the corresponding distribution must be also close to the complete ignorance case. The complete ignorance state reflects the case where we can not take a decision; so this case is not preferable. It is represented in probability distribution as a uniform distribution over the universe set X while in possibility distribution, it is represented as a distribution where all elements are totally possible: the possibility of all the elements of X is equal to 1.
We can find simply that Klir, normalized, symmetric, and VT transformations satisfy the ignorance preservation. In other hand, since the asymmetric transformation does not fulfil the strong preference preservation, the complete ignorance case is not satisfied. It gives different values of possibility for equal probability values.
For a probability distribution where the elements are much closed, we aim to find a transformation, which provides a best distinction between the confused elements. To see the behaviour of the previous transformations in this case, we consider the following example of a probability distribution $p = (0.16, 0.16, 0.16, 0.16, 0.16, 0.2)$ which is very close to the complete ignorance state. The two latest elements are very close which entails a confused case. We will compare the previous transformations in searching the one, which gives the best distinction between these two confused elements. The Fig. I shows the results.

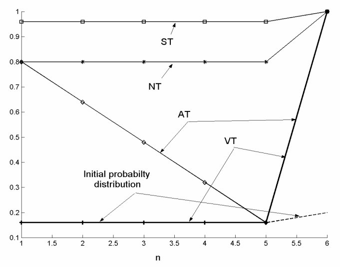

Fig. I. Comparison of the behaviour of the previous transformations in a case close of the complete ignorance state : Variable Transformation (VT), Normalized Transformation (NT), Asymmetric Transformation of Dubois and Prade (AT) and Symmetric Transformation of Dubois and Prade (ST).

We can find that the VT gives the best distinction between the two confused elements. This result is obtained for the maximum value of $k = k_{max}$.

3.2 Specificity criterion

In the most specific possibility distribution, only one element w of the universe X has a value equal to 1 while all the other ones have a value equal to zero. Hence a specificity measure, or Yager like measure of specificity, is evaluated by the degree to which a given distribution is close to the most specific one. Let π_1 and π_2 denote two possibility distributions defined on a universe X, we say that π_2 is more specific than π_1 if (Dubois 93) and (Lasserre 1999):

$$\pi_2 < \pi_1 \quad (10)$$

We call (10) the maximum specificity principle. It allows finding the most informative distribution, which looses the minimum of information. Indeed, we can conclude from (10) that, the most specific distribution is the one which has the smallest fuzzy set, thus the degree of specificity of a distribution π is measured by (Lasserre 1999):

$$S_p = \sum_{i=1}^{n} \pi_i \tag{11}$$

For the most specific distribution, S_p is equal to 1.

Proposition. The possibility distribution which is the most specific one and satisfies the consistency principle, defined by (8), has the specificity degree:

$$S_{max} = 1 + \sum_{i=2}^{n} p_i \tag{12}$$

Proof.

In a possibility distribution, which satisfies the normalization criterion, the possibility of the most frequently element in X is equal to 1. For the other elements, the possibility of each one of them must be at least equal to its value of probability to satisfy the consistency principle, defined by (8), and at most equal to its probability, to maximize the specificity degree. This leads to the maximum specificity degree defined by (12).

We call the difference between the specificity degree S_p and the maximum specificity degree S_{max} the maximum specificity principle. To evaluate the specificity of VT with the one of the previous transformations, we will take the following example. Assume that we have a probability distribution $p = (0, 0.02, 0.03, 0.1, 0.15, 0.4, 0.15, 0.1, 0.03, 0.02, 0)$. The corresponding possibility distributions according to the previous transformations are shown in Fig. II. We can find that the VT, with $k = k_{max}$, is the most specific transformation.

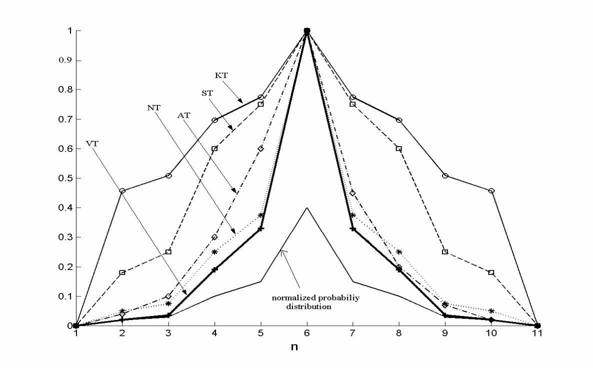

Fig. II. Comparison of the specificity for : Variable Transformation (VT), Normalized Transformation (NT), Asymmetric Transformation (AT), Symmetric Transformation (ST) and Klir Transformation (KT).

4 Conclusion

In this paper, we have explained the most important consistency principles to transform a probability distribution into possibility one or conversely. We have studied the most common transformations constructed in using the previous consistency principles. These transformations are: normalized transformation, Klir transformation, symmetric and asymmetric transformations of Dubois and Prade. We have proposed a transformation from probability to possibility which we have named Variable

Transformation (VT) because it changes its specificity degree with the value of a constant parameter called *k*. This Transformation can be the most specific transformation and the transformation that best distinguishes the confused elements. We have evaluated and compared the behaviour of the previous transformations in using some evaluation criterions.

Our team of research works on diagnosis in using fuzzy classification methods for data coming from industrial and medical sectors. We use the Fuzzy Pattern Matching (FPM) (Mouchaweh et al. 2002b) as a classification method and the symmetric probability-possibility transformation of Dubois and Prade to construct the densities of possibilities. These densities are used to assign each new sample to its suitable class. Each class represents a functioning mode of a system.

Although it is more specific than the symmetric transformation, the asymmetric transformation can not be used in diagnosis by pattern recognition because it does not fulfil the preference and shape preservation criteria.

We have used the Variable Transformation to improve the performances of FPM in excluding the outliers and in detecting the zones of high-density (Mouchaweh et al. 2002a). Indeed, the outliers represent the noise and the zones of high density represent the sub-classes or sub-functioning modes. The outliers have relatively very small probability and the samples of the zones of high density have high probability. Thanks to the high specificity degree of the VT, the patterns belonging to zones of high density will be detected and the outliers will be eliminated. The VT can do that in working as a filter with a variable bandwidth which depends on the value of *k*. In (Mouchaweh et al. 2002a), we explain in detail the use of the VT, in a painting system, to detect the suitable sub-functioning mode corresponding to a high quality of painting, and to predict the evolution of the painting system towards an abnormal functioning mode.

References

Devillez A, Mouchaweh MS and Billaudel P (2004), A process monitoring module based on fuzzy logic and Pattern Recognition, International Journal of Approximate Reasoning 37/1, 43-70.

Dubois D, Prade H (1983). Unfair coins and necessity measures: towards a possibilistic inerpretation of histogram, Fuzzy Sets and Systems 10 ‘1’: 5 –20.

Dubois D, Prade H (1987). *Théorie des possibilités*, Masson.

Dubois D, Prade H (1993). On possibility/probability transformations, Fuzzy Logic : 103-112.

Klir GJ (1992). Probability-possibility transformations: a comparison, Int J General Systems 21: 291-310.

Klir GJ (1993). Information-preserving probability-possibility transformations: recent developments, Fuzzy Logic: 417-428.

Krishnapuram R, Keller JM (1993). A possibilistic approach to clustering, IEEE transactions on fuzzy systems 1 ‘2’: 98-109.

Lasserre V (1999). Modélisation floues des incertitudes de mesures de capteurs, Thèse présentée devant l’Université de Savoie.

Mouchaweh MS, Billaudel P and Lecolier GV (2002a). Elimination du bruit et détection des zones de haute density, Revue d’Intelligence Artificielle 16 ‘3’.

Mouchaweh MS, Devillez A, Lecolier GV and Billaudel P (2002b). Incremental learning in real time using fuzzy pattern matching, accepted paper in Fuzzy Sets and Systems.

Mouchaweh MS (2002). Conception d’un système de diagnostic adaptatif et prédictif basé sur la méthode Fuzzy pattern Matching pour la surveillance en ligne des systèmes évolutifs, Thèse soutenue à l’Université de Reims, France.

Mouchaweh MS (2004), Diagnosis in real time for evolutionary processes using Pattern Recognition and Possibility theory, Invited paper in International Journal of Computational Cognition 2/1, 79-112.

Oussalah M (2000). On the probability/possibility transformations: a comparative analysis, Int. J. General Systems 29 ‘5’: 671-718.

Parsons S, Hunter A (1998). A review of uncertainty handling formalisms, Applications of Uncertainty Formalisms ‘1455’: 8-37.

Shafer G (1976). A Mathematical theory of evidence. Princeton University Press, Princeton.

Walley P (1996). Measures of uncertainty in expert systems, Artificial Intelligence 83 : 1-58.

Zadeh LA (1965). Fuzzy sets, Informations and control 8 : 338-353.

Zadeh LA (1978). Fuzzy sets as a baisis for a theory of possibility, Fuzzy Sets and Systems 1 ‘1’: 3-28.