Expert Systems with Applications 39 (2012) 2279–2287

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Supervised Pseudo Self-Evolving Cerebellar algorithm for generating fuzzy membership functions

K.K. Ang [a], C. Quek [b,*]

[a] Institute for Infocomm Research, A*STAR 1 Fusionopolis Way, #21-01, Connexis 138632, Singapore
[b] Centre for Computational Intelligence, Nanyang Technological University, School of Computer Engineering, Blk N4 #2A-32, Nanyang Avenue 639798, Singapore

ARTICLE INFO

Keywords:
Fuzzy membership function
Linguistic modeling
Interpretability
Classification

ABSTRACT

The proper generation of fuzzy membership function is of fundamental importance in fuzzy applications. The effectiveness of the membership functions in pattern classifications can be objectively measured in terms of interpretability and classification accuracy in the conformity of the decision boundaries to the inherent probabilistic decision boundaries of the training data. This paper presents the Supervised Pseudo Self-Evolving Cerebellar (SPSEC) algorithm that is bio-inspired from the two-stage development process of the human nervous system whereby the basic architecture are first laid out without any activity-dependent processes and then refined in activity-dependent ways. SPSEC first constructs a cerebellar-like structure in which neurons with high trophic factors evolves to form membership functions that relate intimately to the probability distributions of the data and concomitantly reconcile with defined semantic properties of linguistic variables. The experimental result of using SPSEC to generate fuzzy membership functions is reported and compared with a selection of algorithms using a publicly available UCI Sonar dataset to illustrate its effectiveness.

1. Introduction

The generation of membership function is an important step in many applications of fuzzy theory (Medasani, Kim, & Krishnapuram, 1998). Most common membership functions are triangular, trapezoidal, Gaussian and bell-shaped (Mendel, 2001; Zenebe & Norcio, 2009; Zhou & Khotanzad, 2007). Fig. 1(a) and (b) depict trapezoidal and Gaussian membership functions, mathematically described by Eqs. (1) and (2) respectively. Triangular and bell-shaped membership functions can be described by Eq. (1) using parameters such that $\beta = \gamma$ and by Eq. (2) using parameters such that $\alpha = \gamma$ and $\beta = \delta$, respectively.

$$\mu_T(x;\alpha,\beta,\gamma,\delta) = \begin{Bmatrix} 0 & x \leqslant \alpha \text{ or } x \geqslant \delta \\ \frac{x-\alpha}{\beta-\alpha} & \alpha < x < \beta \\ 1 & \beta \leqslant x \leqslant \gamma \\ \frac{\delta-x}{\delta-\gamma} & \gamma < x < \delta \end{Bmatrix}, \quad (1)$$

$$\mu_G(x;\alpha,\beta,\gamma,\delta) = \begin{Bmatrix} e^{\frac{-(x-\beta)^2}{2\alpha^2}} & x < \beta \\ 1 & \beta \leqslant x \leqslant \delta \\ e^{\frac{-(x-\delta)^2}{2\gamma^2}} & x > \delta \end{Bmatrix}, \quad (2)$$

where α, β, γ, δ are parameters of the membership function.

Although fuzzy membership functions can be mathematically described using Eqs. (1) and (2), there is no uniformity in the interpretation of what a fuzzy membership grade means. Three main semantics for membership functions exist in the literature; namely, *similarity*, *preference* and *uncertainty* (Dubois & Prade, 1997). Considering the degree of membership $\mu_F(u)$ of an element u in a fuzzy set F defined on the universe U, the three interpretations of this degree are:

- *Similarity* – is the proximity of u to a prototype element F (Bellman, Kalaba, & Zadeh, 1966). If u is exactly F then $\mu_F(u)$ is 1.
- *Preference* – is the preference in favor of selecting u where F represents a set of preferred objects (Bellman & Zadeh, 1970).
- *Uncertainty* – is the plausibility that a parameter x has value u, given that x is F. This was proposed in (Zadeh, 1978) when possibility theory and approximate reasoning were introduced.

In addition, there are no measures available to evaluate the goodness or correctness of the membership functions generated (Medasani et al., 1998). As a result, there are several approaches to the generation of fuzzy membership functions:

- *Heuristics* – uses predefined membership functions where associated parameters are provided by human experts. This approach yields reasonably smooth membership functions that are easily manipulated by fuzzy operators, but the parameters have to be manually optimized (Medasani et al., 1998).

* Corresponding author.
E-mail addresses: kkang@i2r.a-star.edu.sg (K.K. Ang), ashcquek@ntu.edu.sg (C. Quek).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.08.001

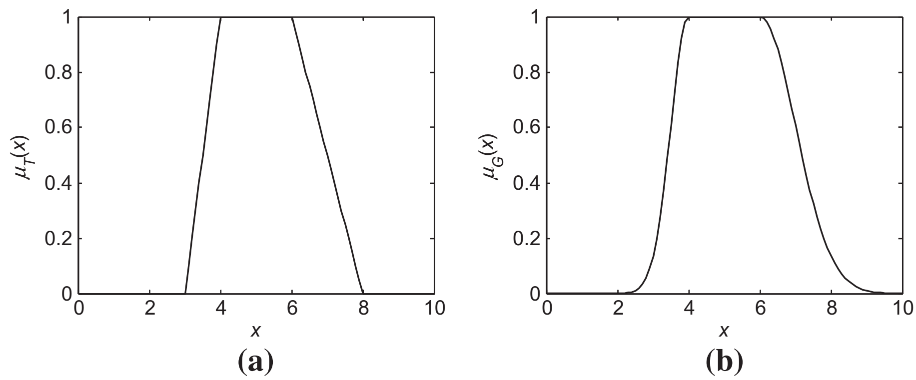

Fig. 1. (a) Trapezoidal membership function $\mu_T(x; 3, 4, 6, 8)$ and (b) Gaussian fuzzy membership functions $\mu_G(x; 0.5, 4, 1, 6)$.

- *Histograms* – provides information for estimating the probability distributions of the data, which can be represented by parameterized functions such as Gaussian, thus directly yielding membership functions. This approach is easy to implement and the membership functions can be used for classifying data (Medasani et al., 1998), but the histograms of different classes frequently overlap, hence limiting the applicability for finding linguistic terms (Duch, Setiono, & Zurada, 2004). Moreover, the randomness represented by probability theory and the vagueness represented by fuzzy set theory are inherently different concepts (Kosko, 1992). Hence, the validation on the interpretation of membership functions generated using this method remains in question (Medasani et al., 1998).
- *Genetic algorithms* – performs a global search for useful and suitable membership functions (Alcala, Alcala-Fdez, & Herrera, 2007; Alcalá-Fdez, Alcalá, Gacto, & Herrera, 2009; Chen, Hong, & Tseng, 2009; Hong, Chen, Wu, & Lee, 2006; Huang, Pasquier, & Quek, 2009). This approach is more complicated as it requires the specification of the number of membership functions, the definition of suitability by an objective function, and various parameters such as how the membership functions are encoded by a chromosome in the Genetic algorithm.
- *Nearest neighbors* – assigns class memberships to a sample instead of a particular class, where the class memberships depend on the sample's distance from its k nearest neighbors (Keller, Gray, & Givens, 1985). This approach is the simplest, but generates membership functions that are not smooth (Medasani et al., 1998).
- *Feedforward neural networks* – employs neurons using sigmoid activation functions that are equivalent to fuzzy membership functions that can be easily extracted (Duch, 2005; Medasani et al., 1998). This approach is capable of generating complex membership functions for classifying data, but the membership values are not necessarily indicative of the similarity of a feature to a class and are unpredictable in regions where there is no training data (Medasani et al., 1998).
- *Clustering* – partitions a dataset into subsets called clusters so that data in each subset are similar. This approach employs unsupervised self-organization which generates membership functions that efficiently covers the regions where numerical data are present (Lin, 1995), but the number of clusters must be known (Medasani et al., 1998) for example *Discrete Clustering Technique* – DCT (Singh, Quek, & Cho, 2008). Although clustering algorithms such as the Robust Agglomerative Gaussian Mixture Decomposition (RAGMD) (Medasani et al., 1998), Adaptive Resonance Theory (ART) (Carpenter, Grossberg, Markuzon, Reynolds, & Rosen, 1992; Lin & Lin, 1997) and Discrete Incremental Clustering (Tung & Quek, 2002a) do not require the specification of the number of clusters, other parameters that affect the number of clusters generated are required. The parameters required are, namely, the retention ratio P in RAGMD (Medasani et al., 1998); the vigilance criterion ρ in ART (Lin & Lin, 1997); and the SLOPE and STEP in DIC (Tung & Quek, 2002a).
- *Probability to possibility transformations* – employs the possibility/probability consistency principle (Zadeh, 1978) to convert probability distribution function to possibility distribution function. This principle expresses a weak connection between possibility and probability where a fuzzy variable is associated with a possibility distribution represented by a membership in the same manner as a random variable is associated with a probability distribution. This approach assumes that the general shapes of the probability functions and possibility functions are similar and thus facilitates the generation of membership functions directly from normalized histograms (Medasani et al., 1998). Therefore, the validation on the interpretation of membership functions generated using this method, which is similar to the histogram method, remains in question (Medasani et al., 1998).

Due to the number of different approaches available in the literature, it is difficult to choose a single approach of generating membership functions that works for most applications (Medasani et al., 1998). However, the clustering method is popularly used to generate membership functions (Nedjah & Mourelle, 2005; Panella & Gallo, 2005; Tung & Quek, 2002b). For example, the Learning Vector Quantization algorithm (Kohonen, 1989) is popularly employed in Mamdani Fuzzy models (Ang, Quek, & Pasquier, 2003; Lin, 1995; Quek & Singh 2005; Zhou & Quek, 1996), and the Fuzzy C-Means algorithm (Bezdek, 1981) is popularly employed in TSK Fuzzy models (Flores-Sintas, Cadenas, & Martin, 1999; Li, Mukaidono, & Turksen, 2002; Panella & Gallo, 2005). Bijective transform (Dubois & Prade, 1983), which is based on the probability to possibility transform approach, has also gained interest in recent year (Masson & Denoeux, 2006). For a review on membership generation techniques, please refer to (Medasani et al., 1998).

Since the main advantage of using fuzzy applications is to abstract humanly interpretable linguistic expressions from available numerical data, the membership functions generated have to reconcile with the semantic properties of a linguistic variable (Casillas, Cordón, Herrera, & Magdalena, 2003). A linguistic variable is formally defined (Zadeh, 1975) with a quintuple (L, $T(L)$, U, G, M) where L is the name of the variable; $T(L)$ is the linguistic term set of L; U is a universe of discourse; G is a syntactic rule that generates $T(L)$; and M is a semantic rule that associates each $T(L)$ with its meaning. Each linguistic term is characterized by a fuzzy set that is described mathematically using a membership function.