

Incorporating Fuzzy Logic in Data Mining Tasks

Lior Rokach

Ben Gurion University, Israel

In this chapter we discuss how fuzzy logic extends the envelop of the main data mining tasks: clustering, classification, regression and association rules. We begin by presenting a formulation of the data mining using fuzzy logic attributes. Then, for each task, we provide a survey of the main algorithms and a detailed description (i.e. pseudo-code) of the most popular algorithms.

INTRODUCTION

There are two main types of uncertainty in supervised learning: statistical and cognitive. Statistical uncertainty deals with the random behavior of nature and all existing data mining techniques can handle the uncertainty that arises (or is assumed to arise) in the natural world from statistical variations or randomness. Cognitive uncertainty, on the other hand, deals with human cognition.

Fuzzy set theory, first introduced by Zadeh in 1965, deals with cognitive uncertainty and seeks to overcome many of the problems found in classical set theory. For example, a major problem faced by researchers of control theory is that a small change in input results in a major change in output. This throws the whole control system into an unstable state. In addition there was also the problem that the representation of subjective knowledge was artificial and inaccurate. Fuzzy set theory is an attempt to confront these difficulties and in this chapter we show how it can be used in data mining tasks.

BACKGROUND

Data mining is a term coined to describe the process of sifting through large and complex databases for identifying valid, novel, useful, and understandable patterns and relationships. Data mining involves the inferring of algorithms that explore the data, develop

the model and discover previously unknown patterns. The model is used for understanding phenomena from the data, analysis and prediction. The accessibility and abundance of data today makes knowledge discovery and data mining a matter of considerable importance and necessity.

We begin by presenting some of the basic concepts of fuzzy logic. The main focus, however, is on those concepts used in the induction process when dealing with data mining. Since fuzzy set theory and fuzzy logic are much broader than the narrow perspective presented here, the interested reader is encouraged to read Zimmermann (2005).

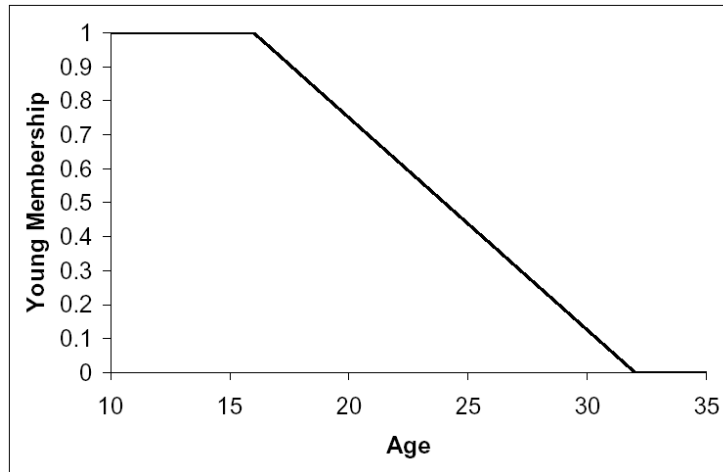
In classical set theory, a certain element either belongs or does not belong to a set. Fuzzy set theory, on the other hand, permits the gradual assessment of the membership of elements in relation to a set.

Let U be a universe of discourse, representing a collection of objects denoted generically by u . A fuzzy set A in a universe of discourse U is characterized by a membership function μ_A which takes values in the interval $[0, 1]$. Where $\mu_A(u) = 0$ means that u is definitely not a member of A and $\mu_A(u) = 1$ means that u is definitely a member of A .

The above definition can be illustrated on the vague set of *Young*. In this case the set U is the set of people. To each person in U , we define the degree of membership to the fuzzy set *Young*. The membership function answers the question "to what degree is person u young?". The easiest way to do this is with a membership function based on the person's age. For example Figure 1 presents the following membership function:

$$\mu_{Young}(u) = \begin{cases} 0 & \text{age}(u) > 32 \\ 1 & \text{age}(u) < 16 \\ \frac{32 - \text{age}(u)}{16} & \text{otherwise} \end{cases} \quad (1)$$

Figure 1. Membership function for the young set

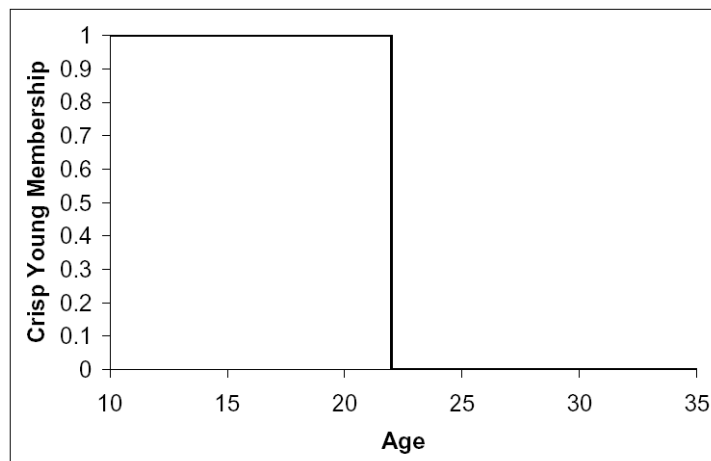


Given this definition, John, who is 18 years old, has degree of youth of 0.875. Philip, 20 years old, has degree of youth of 0.75. Unlike probability theory, degrees of membership do not have to add up to 1 across all objects and therefore either many or few objects in the set may have high membership. However, an object's

membership in a set (such as “young”) and the set's complement (“not young”) must still sum to 1.

The main difference between classical set theory and fuzzy set theory is that the latter admits to partial set membership. A classical or crisp set, then, is a fuzzy set that restricts its membership values to $\{0,1\}$, the

Figure 2. Membership function for the crisp young set



6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/incorporating-fuzzy-logic-data-mining/10348

Related Content

Hardware and Software Cyber Security Tools

Tamalika Das, Nabonita Nath, Kshounish Acharyya, Shirsa Chakraborty and Parag Chatterjee (2023). *AI Tools for Protecting and Preventing Sophisticated Cyber Attacks* (pp. 109-131).

www.irma-international.org/chapter/hardware-and-software-cyber-security-tools/328596

News-Seekers vs. Gate-Keepers: How Audiences and Newsrooms Prioritize Stories in Print and Online Content

Sharon E. Jarvis and Maegan Stephens (2015). *International Journal of Signs and Semiotic Systems* (pp. 50-63).

www.irma-international.org/article/news-seekers-vs-gate-keepers/142500

Data Security With Replication for Cloud Computing

Cigdem Bakir and Mecit Yuzkat (2025). *Artificial Intelligence for Cloud-Native Software Engineering* (pp. 357-376).

www.irma-international.org/chapter/data-security-with-replication-for-cloud-computing/378783

Text Clustering Using PSO Based Dynamic Adaptive SOM for Detecting Emergent Trends

Chandrakala D, Sumathi S, Saran Kumar A and Sathish J (2019). *International Journal of Intelligent Information Technologies* (pp. 64-78).

www.irma-international.org/article/text-clustering-using-pso-based-dynamic-adaptive-som-for-detecting-emergent-trends/230877

Symbolic Data Analysis: A Paradigm for Complex Data Mining?

Sandra Elizabeth González Císaro and Héctor Oscar Nigro (2014). *International Journal of Signs and Semiotic Systems* (pp. 1-9).

www.irma-international.org/article/symbolic-data-analysis/104639