

## MỘT PHƯƠNG PHÁP ĐỊNH LƯỢNG GIÁ TRỊ NGÔN NGỮ CHO TẬP MẪU HUẤN LUYỆN TRONG ĐIỀU KIỆN HẠN CHẾ

Lê Văn Tường Lân

Khoa Công nghệ Thông tin, Trường Đại học Khoa học, Đại học Huế

Email: lvtlan@yahoo.com

Ngày nhận bài: 01/7/2019; ngày hoàn thành phần biên: 02/7/2019; ngày duyệt đăng: 02/7/2019

### TÓM TẮT

Tập mẫu đóng vai trò quan trọng trong quá trình huấn luyện. Khi miền trị của các thuộc tính trong tập mẫu huấn luyện là chưa thuần nhất, việc làm thuần nhất tập huấn luyện là bắt buộc. Đại số gia tử là một công cụ hữu ích để làm thuần nhất tập huấn luyện, bằng cách chuyển miền dữ liệu của thuộc tính chưa thuần nhất thành miền dữ liệu chứa các giá trị ngôn ngữ hay định lượng các giá trị ngôn ngữ về các giá trị kinh điển. Trong quá trình thuần nhất, cần phải biết các giá trị  $\psi_{min}$ ,  $\psi_{max}$  của miền trị kinh điển, tuy vậy trong thực tế, nhiều lúc ta chưa biết cụ thể giá trị  $\psi_{min}$ ,  $\psi_{max}$  của thuộc tính đang xét. Trong bài báo này, chúng ta xây dựng một cách thức để có thể định lượng các giá trị ngôn ngữ khi không biết miền giá trị  $[\psi_{min}, \psi_{max}]$  mà chỉ biết đoạn con  $[\psi_1, \psi_2]$  của chúng.

**Từ khoá:** Tập mẫu huấn luyện, Giá trị ngôn ngữ, Cây quyết định mờ.

## A METHOD TO DETERMINE THE LINGUISTIC VALUES IN THE LIMITED CONDITIONS OF TRAINING DATA SET

Le Van Tuong Lan

Faculty of Information Technology, University of Sciences, Hue University

Email: lvtlan@yahoo.com

### ABSTRACT

Sample training data set plays an important role in the training process. When the value of the attribute domain may be value or linguistics, we need a method to homogenise sample training data set. Hedge algebra is a useful tool to make the training set homogeneous by changing the values of mixed domain to homogeneous data domain that only contains linguistics or values. In the process of homogeneous data domain, we have to know the values  $\psi_{min}$ ,  $\psi_{max}$ . However, in reality, we do not know the values  $\psi_{min}$ ,  $\psi_{max}$  exactly. In this paper, we present a method to determine the linguistic values when we only know the sub values  $[\psi_1, \psi_2]$  without knowing the values  $[\psi_{min}, \psi_{max}]$  exactly.

**Keywords:** Fuzzy decision tree, linguistic values, training data set.



**Lê Văn Tường Lâm** sinh năm 1974 tại thành phố Huế. Ông tốt nghiệp cử nhân chuyên ngành Toán – Tin học tại Trường Đại học Khoa học, Đại học Huế năm 1996 và thạc sĩ chuyên ngành Công nghệ thông tin tại Trường Đại học Bách khoa Hà Nội, năm 2002. Hiện đang là Nghiên cứu sinh tại Trường Đại học Khoa học, Đại học Huế, chuyên ngành Khoa học máy tính. Ông công tác tại khoa Công nghệ thông tin, Trường Đại học Khoa học, Đại học Huế từ năm 1996 đến nay.

*Lĩnh vực nghiên cứu:* Khai phá dữ liệu, công nghệ phần mềm.