



Lý thuyết đại số gia tử và ứng dụng trong các bài toán trích rút tri thức

Trần Xuân Thanh^{1*}

¹ Khoa Công nghệ thông tin, Trường Đại học Công nghệ Đông Á

*Email: thanhtx@eaut.edu.vn

Tóm tắt

Trong bối cảnh dữ liệu bùng nổ ngày càng phức tạp và không chắc chắn, nhu cầu về các mô hình toán học có khả năng xử lý thông tin mơ hồ và trích rút tri thức từ dữ liệu ngày càng quan trọng. Lý thuyết Đại số gia tử (Hedge Algebra - ĐSGT) là mô hình toán học dựa trên lý thuyết tập mờ, cho phép biểu diễn và xử lý các hạng từ ngôn ngữ một cách có hệ thống. Bài báo này trình bày nền tảng lý thuyết của ĐSGT, trong đó nhấn mạnh các đặc tính toán học, cấu trúc đại số và phương pháp định lượng ngữ nghĩa của các hạng từ. Đồng thời giới thiệu một công trình điển hình có kết quả tốt khi ứng dụng lý thuyết ĐSGT trong bài toán trích rút tri thức, đặc biệt là trong xử lý ngôn ngữ tự nhiên, hệ hỗ trợ ra quyết định và phân loại dữ liệu không chắc chắn. Các kết quả có được cho thấy ĐSGT giúp cải thiện khả năng diễn giải tập mờ, đồng thời mang đến một hướng tiếp cận khả thi trong các hệ thống thông minh. Cuối cùng, bài báo cũng phân tích các xu hướng nghiên cứu hiện nay và định hướng phát triển tiếp theo của ĐSGT trong lĩnh vực khoa học dữ liệu và trí tuệ nhân tạo.

Từ khóa: Đại số gia tử, tập mờ, khai phá dữ liệu, trích rút tri thức, trí tuệ nhân tạo, xử lý dữ liệu mờ

Abstract

In the context of increasingly complex and uncertain data explosion, the demand for mathematical models capable of handling ambiguous information and extracting knowledge from data has become ever more critical. Hedge Algebra (HA) is a broad mathematical framework generalizing fuzzy set theory, which enables the systematic representation and processing of linguistic terms. This paper presents the theoretical foundations of Hedge Algebra, with particular emphasis on its mathematical properties, algebraic structure, and semantic quantification methods for linguistic terms. It further introduces a representative work demonstrating promising results in applying HA theory to knowledge extraction tasks, especially in natural language processing, decision support systems, and uncertain data classification. The obtained results indicate that HA enhances the interpretability and modeling capability of fuzzy data, while offering a viable approach for intelligent systems. Finally, the paper analyzes current research trends and outlines prospective directions for the further development of Hedge Algebra in the fields of data science and artificial intelligence.

Keywords: Hedge Algebra, fuzzy sets, data mining, knowledge extraction, artificial intelligence, fuzzy data processing



1. Mở đầu

Trong thời đại dữ liệu lớn bùng nổ, việc khai phá và trích rút tri thức từ nguồn dữ liệu lớn, phức tạp và không chắc chắn ngày càng trở nên thách thức. Các phương pháp truyền thống trong khai phá dữ liệu thường dựa trên lý thuyết xác suất thống kê và lý thuyết logic mờ gặp khó khăn khi xử lý thông tin mang tính mơ hồ hoặc biểu diễn tri thức dạng ngôn ngữ tự nhiên. Để giải quyết các thách thức này, Lý thuyết ĐSGT được đề xuất như một mô hình toán học mạnh mẽ giúp mô hình hóa và xử lý dữ liệu không chắc chắn thông qua cấu trúc đại số của các hạng từ ngôn ngữ.

Trong bài báo này, chúng tôi trình bày nền tảng lý thuyết của ĐSGT, tập trung vào những khái niệm, định nghĩa khái quát và các phương pháp định lượng ngữ nghĩa. Đồng thời, bài báo cũng giới thiệu các ứng dụng trong các bài toán thực tế của ĐSGT trong việc trích rút tri thức, đặc biệt là trong việc xử lý dữ liệu mờ và hỗ trợ ra quyết định. Bài báo này sẽ cung cấp một góc nhìn tổng quan về ĐSGT và những ứng dụng tiềm năng mà nó mang lại, từ đó góp phần thúc đẩy nghiên cứu và ứng dụng lý thuyết này vào các hệ thống khai thác tri thức trong thực tế.

2. Lý thuyết đại số gia tử

Vào đầu những năm 1990 Nguyễn Cát Hồ cùng Wechler đã giới thiệu lần đầu tiên lý thuyết ĐSGT trong [16]. Lý thuyết ĐSGT cho phép thao tác trực tiếp trên ngôn ngữ tự nhiên trên cơ sở mô tả miền xác định biến ngôn ngữ dưới dạng đại số gia tử. Nó đã được ứng dụng nhiều trong các bài toán khác nhau như bài toán khai phá dữ liệu [9, 10, 11], điều khiển mờ [12], xử lý hình ảnh và thị giác máy tính [13],... ĐSGT đã cung cấp mô hình toán học giúp cho việc tính toán ngữ nghĩa dựa trên tập mờ từ ngữ nghĩa vốn có tự nhiên của các từ ngôn ngữ. Phương pháp này dựa trên tính chất có thứ tự ngữ nghĩa của các từ trong miền giá trị ngôn ngữ trong biến ngôn ngữ. Từ đó đề xuất một số mô hình toán học nhằm sinh ngữ nghĩa có thể thực hiện tính toán trên tập mờ từ ngữ nghĩa định tính sẵn có của ngôn ngữ tự nhiên.

Khi xét giá trị thuộc tính A của cơ sở dữ liệu với hai miền giá trị: Giá trị số trong miền giá trị U_A (A là biến số) và hạng từ ngôn ngữ trong miền ngôn ngữ LD_A (A là biến ngôn ngữ). Trong thực tế hai miền giá trị tồn tại song hành, chúng có mối quan hệ chặt chẽ với nhau. Do đó, khi U_A có chứa cấu trúc toán học thì LD_A cũng có một cấu trúc toán học. Lý thuyết ĐSGT nghiên cứu cấu trúc đại số dựa trên quan hệ thứ tự ngữ nghĩa vốn có của các hạng từ trong LD_A . Ví dụ: Trong miền hạng từ của biến AGE (tuổi), dựa trên giá trị ngữ nghĩa vốn có trong thực tế, luôn tồn tại một quan hệ có thứ tự về mặt ngữ nghĩa là 'very young' \leq 'young' \leq 'middle' \leq 'little old' \leq 'very old'.

Lý thuyết ĐSGT nghiên cứu về cấu trúc ngữ nghĩa của miền hạng từ của các biến ngôn ngữ dựa trên ngữ nghĩa vốn có của các hạng từ trong ngôn ngữ tự nhiên. Phương pháp tiên đề đã được sử dụng cho Lý thuyết ĐSGT và định lượng ngữ nghĩa trong ĐSGT dựa trên tính chất ngữ nghĩa vốn có trong thế giới thực của các hạng từ và gia từ ngôn ngữ. Điều này tạo cơ sở giúp hình thức hóa mạnh để tính toán các đại lượng định lượng khác nhau của các hạng từ như khoảng tính mờ, độ đo mờ, giá trị định lượng. Từ giá trị định lượng ngữ nghĩa của hạng từ, xây dựng được các thủ tục để sinh ngữ nghĩa dựa trên tập mờ của các hạng từ trong việc giải quyết các bài toán thực tế. Lý thuyết ĐSGT được sử dụng để giải quyết nhiều lớp bài toán khác nhau như điều khiển mờ [20-21]; phân lớp và hồi quy mờ [18, 19, 22]; dự báo chuỗi thời gian [23, 24, 25]; robotics [26, 27]; ... đạt nhiều kết quả tốt so với các phương pháp đã sử dụng trước đó.

2.1. Cấu trúc đại số được xây dựng dựa trên quan hệ thứ tự về mặt ngữ nghĩa

Dựa vào thứ tự về mặt ngữ nghĩa khi quan sát và phân tích trong thực tế của các hạng từ trên miền LD_A của biến ngôn ngữ A , LD_A được coi như một cấu trúc đại số dựa trên thứ tự ngữ nghĩa, ký hiệu là $\wp_A = (X, G, C, H, \leq)$.

Trong đó:

- $X = LD_A$ là miền hạng từ của các biến ngôn ngữ A ;
- $G = \{c^-, c^+\}$ là các phần tử sinh hoặc hạng từ nguyên tử;
- C là tập các hằng ngôn ngữ, $\mathbf{0}$ tương ứng với hạng từ có ngữ nghĩa nhỏ nhất, $\mathbf{1}$ tương ứng với hạng từ có ngữ nghĩa lớn nhất, \mathbf{W} tương ứng với hạng từ có ngữ nghĩa trung bình;
- H là tập các gia từ ngôn ngữ;
- \leq là ký hiệu cho quan hệ thứ tự ngữ nghĩa trong X .

Các hạng từ trong cấu trúc \wp_A có dạng h_n, \dots, h_{1c} , với $c \in \{c^-, c^+\}$ và $h_j \in H, j=1, \dots, n$.

Xét ví dụ miền hạng từ của biến ngôn ngữ AGE trong ngôn ngữ tự nhiên tạo thành một cấu trúc đại số và các thành phần như sau:

- $G = \{\text{'young'}, \text{'old'}\}$, 'young' là phần tử sinh âm, 'old' là phần tử sinh dương. Đây là hai phần tử sinh trong cấu trúc \wp_{AGE}
- Tập các gia từ ngôn ngữ H gồm các từ như 'very' (V), 'more' (M), 'extremely' (E), 'rather' (R), 'little' (L). Các gia từ ngôn ngữ là các từ có thể thay đổi ngữ nghĩa của hạng từ mà nó tác động để tạo ra một hạng từ mới. Hạng từ mới và hạng từ ban đầu luôn so sánh được với nhau về thứ tự ngữ nghĩa.

Ví dụ: ‘very old’ \geq ‘old’, ‘old’ \geq ‘little old’, ‘young’ \geq ‘very young’. Các gia tử ngôn ngữ được coi như các toán tử một ngôi của \wp_{AGE}

- Tồn tại một quan hệ thứ tự \leq trong X sản sinh từ ngữ nghĩa vốn có của các hạng từ. Chẳng hạn: ‘young’ \leq ‘old’ \leq ‘very old’ hoặc ‘very very old’ \geq ‘very old’ \geq ‘old’. Quan hệ \leq được gọi là quan hệ thứ tự về ngữ nghĩa.

- C là tập các hạng từ như ‘absolutely young’, ‘middle age’, ‘absolutely old’ mà ngữ nghĩa của các hạng từ không thay đổi khi tác động gia tử ngôn ngữ. Các hạng từ này được coi là các hằng ngôn ngữ trong X , ký hiệu lần lượt là θ , W , I . Các hạng từ này luôn có thứ tự là $\theta \leq W \leq I$.

2.2. Tính chất ngữ nghĩa của cấu trúc dựa trên quan hệ thứ tự

a) Dấu đại số của gia tử và các hạng từ của \wp_A

Các hạng từ nguyên thủy c^- và c^+ có xu hướng ngữ nghĩa khác nhau được xác định bởi $Vc^- \leq c^+$, nhưng $c^+ \leq Vc^+$. Ví dụ: ‘very young’ \leq ‘young’, ‘old’ \leq ‘very old’. Do đó, với mọi biến ngôn ngữ A , dấu của hai hạng từ nguyên thủy được xác định là $sign(c^-) = -1$ và $sign(c^+) = +1$. Với gia tử ‘rather’, ta có ‘rather young’ \geq ‘young’, nhưng ‘rather old’ \leq ‘old’. Dựa trên quan hệ thứ tự của các từ ngôn ngữ mới sinh ra khi sử dụng các gia tử so với từ ban đầu thì ‘rather’ được coi như gia tử âm và $sign(R) = -1$, ‘very’ được coi là gia tử dương và $sign(V) = +1$. Ký hiệu Tập hợp H bao gồm các gia tử âm, trong khi H^+ chứa các gia tử dương. Trong ví dụ trên về biến ngôn ngữ AGE, ta có: $H^+ = \{M, V, E\}$ và $H = \{R, L\}$.

b) Quan hệ dấu của các hạng từ và dấu của các gia tử

Ngữ nghĩa dựa trên thứ tự của các từ chỉ ra rằng bất cứ gia tử h nào cũng đều có thể tác động đến ảnh hưởng của gia tử h trong hạng từ hx , $x \in X$, để tạo ra sự thay đổi xu hướng ngữ nghĩa của hx , có thể là cùng xu hướng (dương) hoặc đảo ngược xu hướng (âm) của gia tử h khi tác động vào hạng từ x . Nếu cùng xu hướng thì dấu của h' đối với h là dấu dương và ký hiệu là $sign(h', h) = +1$. Nếu ngược xu hướng thì dấu của h' đối với h là dấu âm và ký hiệu là $sign(h', h) = -1$.

Ví dụ: $h' = \text{‘extremely’}$ (E), $h'' = \text{‘rather’}$ (R) và $h = \text{‘very’}$ (V), ta có $sign(E, V) = +1$ vì ‘young’ \geq ‘V young’ \geq ‘EV young’, $sign(R, V) = -1$ vì ‘young’ \geq ‘RV young’ \geq ‘V young’. Với $H^+ = \{M, V, E\}$ và $H = \{R, L\}$, bảng dấu giữa các gia tử dựa trên xu hướng thay đổi ngữ nghĩa trong ngôn ngữ tự nhiên được cho như Bảng 1.

Bảng 1: Quan hệ về dấu giữa một gia tử trên dòng với gia tử trên cột

$sign(k, h)$	E	V	M	R	L
E	+	+	+	-	+
V	+	+	+	-	+
M	+	+	+	-	+
R	-	-	-	+	-
L	-	-	-	+	-

Với mọi hạng từ $x \in X$ có dạng $x = h_n \dots h_1 c$, $c \in G$, dấu của x được xác định bởi biểu thức $sign(x) = sign(h_n, h_{n-1}) \dots sign(h_2, h_1) sign(h_1) sign(c) \in \{-1, +1\}$.

Ý nghĩa của dấu hạng từ được thể hiện như sau:

$$sign(hx) = -1 \Leftrightarrow hx \leq x \text{ và } sign(hx) = +1 \Leftrightarrow hx \geq x \quad (1)$$

2.3. Định lượng ngữ nghĩa của từ ngôn ngữ trong ĐSGT

a) Ánh xạ định lượng ngữ nghĩa của từ ngôn ngữ

Những hạng từ trong miền ngôn ngữ của các biến ngôn ngữ A có thứ tự dựa trên ngữ nghĩa vốn có, do đó giá trị định lượng ngữ nghĩa ngôn ngữ của mỗi hạng từ trong miền tham chiếu $[0, 1]$ (miền tham chiếu chuẩn hoá cho mọi biến ngôn ngữ A) cần đảm bảo duy trì quan hệ thứ tự giữa các hạng từ.

Định nghĩa 1: Một ánh xạ $f: X \rightarrow [0, 1]$, gán mỗi hạng từ trong X bởi một giá trị trong khoảng $[0, 1]$, được gọi là ánh xạ định lượng ngữ nghĩa (a semantically quantifying mapping – SQM) khi thỏa các điều kiện sau đây:

- i. f là ánh xạ một – một và $f(X)$ trù mật trong $[0, 1]$, tức là một giá trị trong $[0, 1]$ được xấp xỉ với ngữ nghĩa giá trị số của một hạng từ nào đó trong X , với mức độ mờ cho trước.
- ii. f bảo toàn thứ tự ngữ nghĩa trong X , tức là $\forall x, y \in X, x \leq y$ thì $f(x) \leq f(y)$.

Cho một ánh xạ định của lượng ngữ nghĩa $f: X \rightarrow [0, 1]$, ký hiệu $I(x)$ là khoảng con nhỏ nhất của $[0, 1]$ chứa mọi ảnh của $f(H(x))$, khoảng này là khoảng đóng bên trái và mở bên phải, ngoại trừ bên phải là 1. Từ các tính chất trên, ta có khẳng định sau:

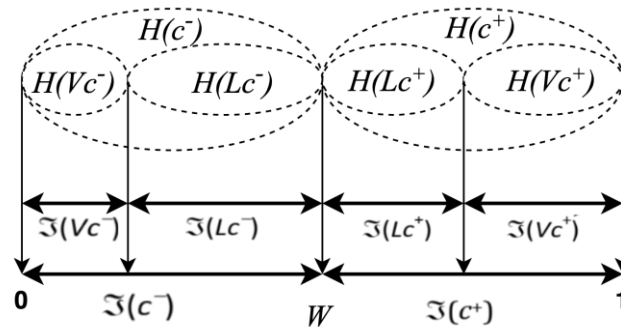
- i. Họ $\{I(hx): h \in H\}$ là phân hoạch của $I(x)$, $\forall x \in X$
- ii. Họ $\{I(x): x \in X_k\}$ là phân hoạch của $[0, 1]$. Khi $k = 1$, $\{I(c^-), I(c^+)\}$ là phân hoạch trên đoạn $[0, 1]$

iii. Họ $\{I(y): y \in X_{k+1}\}$ tạo phân hoạch mịn hơn $\{I(x): x \in X_k\}$, và ta có

$$I(x) = U \{I(y): y \in X_{k+1} \& y = hx \& h \in H\}$$

b) Độ đo tính mờ của hạng từ và gia tử

Vì $H(x)$ gồm các hạng từ mà vẫn chứa ngữ nghĩa vốn có của x nên $H(x)$ được coi là mô hình mờ của x . Do đó, khoảng $I(x)$ được gọi là khoảng tính mờ của x và độ dài của $I(x)$, $|I(x)|$, được hiểu là độ đo tính mờ của x . Ký hiệu $fm(x) = |I(x)|$, với mọi $x \in H(G)$. Xét về ngữ nghĩa, $I(x)$ chứa các giá trị số mà chúng tương thích với ngữ nghĩa của x ở mức độ được xác định bởi $k = |x|$, nên $I(x)$ được gọi là k -khoảng tính mờ của x . Ví dụ trong Hình 1 dưới đây minh họa cho khoảng tính mờ của các hạng từ trong tập $X_{(2)}$ của một cấu trúc ĐSGT với tập gia tử $H = \{L, V\}$.



Hình 1: [32] Các khoảng tính mờ của các hạng từ trong $X_{(2)}$ với tập gia tử

$$H = \{L, V\}$$

c) Khoảng tính mờ

Khoảng giá trị này được gọi là *khoảng tính mờ* (ngữ nghĩa khoảng giá trị) của hạng từ. Ví dụ, xét biến ngôn ngữ ‘AGE’, hạng từ ‘young’ được liên kết với khoảng giá trị $[0, 45]$ trong miền tham chiếu $[0, 75]$ của biến ‘AGE’ trong cộng đồng, với 75 tuổi là tuổi thọ trung bình. Trong lý thuyết ĐSGT, ngữ nghĩa khoảng của các hạng từ trong một cấu trúc ĐSGT của một biến ngôn ngữ trên miền tham chiếu $U = [0, 1]$ được xác định bởi ánh xạ f . Tập chứa tất cả các hạng từ vẫn còn kế thừa ngữ nghĩa vốn có của hạng từ x , ảnh $f(H(x))$ qua ánh xạ đẳng cấu thứ tự f nên tập trung vào một khoảng con của $[0, 1]$, ký hiệu khoảng con này là $Z(x)$. Vì $Z(x)$ là ảnh đẳng cấu của mô hình mờ của x nên được gọi là khoảng tính mờ của x . Ví dụ, tuổi của những người trong khoảng $[0, 45]$ như trong ví dụ trên thì chúng ta coi những người đó có tuổi là ‘young’. Số lượng giá trị trong khoảng giá trị này hay kích thước của khoảng này biểu diễn tính mờ của hạng từ ‘young’. Rõ ràng, khoảng $Z(x)$ càng lớn, hạng từ x càng mờ. Do đó, độ dài của $Z(x)$ được gọi là độ đo tính mờ của hạng từ x , ký hiệu bởi $fm(x)$. Tức là $fm(x) = |Z(x)|$.

3. Ứng dụng của ĐSGT

ĐSGT là một mô hình toán học mạnh mẽ giúp mở rộng khả năng biểu diễn và xử lý thông tin mờ. Với sự kết hợp của lý thuyết tập mờ và biến ngôn ngữ, ĐSGT cho phép mô hình hóa các yếu tố không chắc chắn và hỗ trợ quá trình ra quyết định trong nhiều lĩnh vực khác nhau.

3.1. Trích rút thông tin và khai phá dữ liệu

Trong lĩnh vực khai phá dữ liệu, việc trích rút thông tin mang tính ngữ nghĩa từ các tập dữ liệu lớn là một bài toán quan trọng và đang nhận được nhiều sự quan tâm [1][2]. Đại số gia tử (ĐSGT) được xem là một công cụ hiệu quả trong việc biểu diễn tri thức dưới dạng ngôn ngữ tự nhiên, giúp cải thiện khả năng hiểu và xử lý dữ liệu so với các biểu diễn thuần số học [16]. Cụ thể, trong bài toán phân lớp (classification), ĐSGT cho phép xây dựng các luật phân lớp có thể hiểu được dưới dạng ngôn ngữ như "nếu tuổi là *trẻ* và thu nhập là *thấp* thì khả năng mua hàng là *thấp*". Điều này không chỉ giúp mô hình dễ giải thích mà còn hỗ trợ ra quyết định trong các hệ thống chuyên gia. Trong bài toán hồi quy (regression), ĐSGT có thể sử dụng để ánh xạ các giá trị đầu vào mờ đến các đầu ra định lượng thông qua các hàm ngữ nghĩa, từ đó tạo ra các mô hình hồi quy mềm giúp dự đoán linh hoạt hơn trong môi trường dữ liệu không chắc chắn hoặc nhiễu. Trong bài toán trích rút câu tóm tắt (extractive summarization), ĐSGT hỗ trợ việc ánh xạ các đặc trưng thống kê (ví dụ: tần suất từ, vị trí câu, độ dài, v.v.) sang các giá trị ngôn ngữ như *cao*, *trung bình*, *thấp*. Sau đó, các giá trị này được xử lý thông qua hệ suy diễn mờ để đánh giá mức độ quan trọng của mỗi câu trong văn bản. Việc này giúp lựa chọn các câu phù hợp nhất để đưa vào bản tóm tắt, đồng thời vẫn giữ được tính trực quan và giải thích được kết quả [4][9][17]. Ví dụ, trong một hệ thống đánh giá sản phẩm, thay vì chỉ sử dụng các điểm số từ 1 đến 5, ĐSGT có thể biểu diễn mức độ hài lòng của khách hàng dưới dạng "rất tốt", "tốt", "trung bình", "kém", và gán các giá trị tương ứng. Điều này giúp phân tích dữ liệu người dùng trở nên tự nhiên, dễ hiểu và phù hợp hơn với hành vi thực tế của con người.

3.2. Hệ hỗ trợ ra quyết định

Các hệ hỗ trợ ra quyết định thường phải xử lý các dữ liệu không chắc chắn hoặc thông tin mơ hồ từ nhiều nguồn khác nhau [3, 4]. ĐSGT giúp mô hình hóa và tổng hợp các đánh giá từ các chuyên gia hoặc dữ liệu thực tế để đưa ra quyết định tối ưu [3, 4]. Ví dụ, trong tuyển dụng nhân sự, các tiêu chí đánh giá như "*kinh nghiệm làm việc*", "*kỹ năng giao tiếp*", "*khả năng sáng tạo*" có thể được biểu diễn bằng các giá trị ngôn ngữ và được tính toán bằng ĐSGT để xếp hạng các ứng viên một cách hiệu quả.

3.3. Ứng dụng trong y học

Trong y học, ĐSGT giúp mô hình hóa các thông tin mơ hồ từ dữ liệu bệnh án, hỗ trợ bác sĩ trong việc chẩn đoán và ra quyết định điều trị [6, 7]. Ví dụ, khi phân tích mức độ nguy hiểm của một bệnh nhân bị tiểu đường, bác sĩ có thể sử dụng các mức đánh giá như "*nguy cơ thấp*", "*nguy cơ trung bình*", "*nguy cơ cao*" thay vì chỉ dựa vào các con số cụ thể như chỉ số đường huyết [8, 9].

3.4. Ứng dụng trong kinh tế và tài chính

Các mô hình tài chính thường xuyên phải xử lý thông tin không chắc chắn, chẳng hạn như dự báo giá cổ phiếu hoặc bài toán đánh giá mức độ rủi ro tín dụng [14]. ĐSGT giúp biểu diễn thông tin mờ về xu hướng thị trường, từ đó hỗ trợ các nhà đầu tư ra quyết định hiệu quả hơn [15]. Ví dụ, thay vì sử dụng các mô hình để dự báo cứng nhắc dựa trên dữ liệu lịch sử, các chuyên gia có thể ứng dụng ĐSGT để thực hiện mô hình hóa những yếu tố tác động như "*lạm phát cao*", "*tăng trưởng kinh tế chậm*" nhằm đưa ra những phân tích linh hoạt hơn [23, 24, 25].

3.5. Điều khiển mờ và tự động hóa

Trong lĩnh vực điều khiển, ĐSGT được ứng dụng để cải thiện một số hệ thống điều khiển mờ, giúp các hệ thống tự động đưa ra quyết định hợp lý hơn [29, 30, 31]. Ví dụ, trong điều khiển nhiệt độ thông minh, hệ thống có thể điều chỉnh mức làm mát hoặc sưởi ấm dựa trên các mô tả như "*hơi lạnh*", "*ấm áp*" thay vì chỉ dựa trên các mức nhiệt độ cố định.

4. Trao đổi và kết luận

Lý thuyết ĐSGT cung cấp một mô hình toán học mạnh mẽ giúp mô hình hóa và xử lý dữ liệu không chắc chắn một cách có hệ thống. Với khả năng định lượng ngữ nghĩa của các hạng từ và xây dựng các phép toán trên miền giá trị của biến ngôn ngữ, ĐSGT đã chứng minh tiềm năng trong nhiều bài toán thực tế. Xu hướng nghiên cứu hiện nay trong khoa học dữ liệu và trí tuệ nhân tạo tập trung vào việc phát triển các mô hình lai ghép giữa toán học và học máy nhằm nâng cao hiệu quả của việc xử lý dữ liệu. ĐSGT có thể đóng vai trò quan trọng trong việc xây dựng các hệ thống thông minh có khả năng diễn giải, đặc biệt là khi kết hợp với các phương pháp học sâu (Deep Learning) hoặc mô hình ngôn ngữ lớn (LLMs). Trong tương lai, nghiên cứu có thể mở rộng theo hướng ứng dụng ĐSGT vào học máy giải thích được (Explainable AI - XAI), tối ưu hóa khai phá dữ liệu. Với những tiềm năng đã được chứng minh, ĐSGT không chỉ là một lý thuyết toán học mà còn là một công cụ hữu ích trong việc phát triển các phương pháp trích rút tri thức từ dữ liệu phức tạp, mở ra nhiều hướng nghiên cứu mới trong kỷ nguyên của dữ liệu lớn và trí tuệ nhân tạo.



Lời cảm ơn

Thay mặt nhóm tác giả, tôi xin gửi lời cảm ơn đến các đồng nghiệp của tôi trong nhóm nghiên cứu ĐSGT. Tôi cũng xin gửi lời cảm ơn sâu sắc đến Tiến sĩ Phạm Đình Phong và Tiến Sĩ Phạm Thị Lan đã hỗ trợ, góp ý cho tôi hoàn thành thành nghiên cứu này.

Tài liệu tham khảo

- [1]. S.Mitra, S.K.Pal, "Data mining in soft computing framework: a survey," *IEEE transactions on neural networks*, vol. 13, no. 1, pp. 3-14, 2002.
- [2]. E. Hüllermeier, "Fuzzy sets in machine learning and data mining," *Applied Soft Computing*, vol. 11, no. 2, pp. 1493-1505, 2011.
- [3]. R.R.Yager, K.M.Ford, and A.J.Cañas, "An approach to the linguistic summarization of data," in *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 1990, pp. 456-468: Springer.
- [4]. J.Kacprzyk, "Intelligent data analysis via linguistic data summaries: a fuzzy logic approach," in *Classification and Information Processing at the Turn of the Millennium*: Springer, 2000, pp. 153-161.
- [5]. J.Kacprzyk and S.Szadrozny, "Towards human consistent data driven decision support systems using verbalization of data mining results via linguistic data summaries," *Bulletin of the Polish Academy of Sciences Technical Sciences*, vol. 58, no. 3, pp. 359-370, 2010.
- [6]. R.M.Catillo-Ortega, N.Marín, and D.Sánchez, "A Fuzzy Approach to the Linguistic Summarization of Time Series," *Journal of Multiple-Valued Logic & Soft Computing*, vol. 17, 2011.
- [7]. G.Moyse, M.J.Lesot, and B.Bouchon-Meunier, "Linguistic summaries for periodicity detection based on mathematical morphology," in *2013 IEEE Symposium on Foundations of Computational Intelligence (FOCI)*, 2013, pp. 106-113: IEEE.
- [8]. G.Moyse, M.J.Lesot, and B. Bouchon-Meunier, "Mathematical morphology tools to evaluate periodic linguistic summaries," in *International Conference on Flexible Query Answering Systems*, 2013, pp. 257-268: Springer.
- [9]. G.Moyse, M.J.Lesot, and B. Bouchon-Meunier, "Oppositions in fuzzy linguistic summaries," in *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on*, 2015, pp. 1-8: IEEE.



- [10]. A.Wilbik and J.Kacprzyk, "A multi-criteria evaluation of linguistic summaries of time series via a measure of informativeness," in *International Conference on Artificial Intelligence and Soft Computing*, 2010, pp. 105- 113: Springer.
- [11]. A.Wilbik, J.M.Keller, and G. L. Alexander, "Linguistic summarization of sensor data for eldercare," in *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, 2011, pp. 2595-2599: IEEE.
- [12]. A.Wilbik, J.M.Keller, and J.C.Bezdek, "Generation of prototypes from sets of linguistic summaries," in *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, 2012, pp. 1-8: IEEE.
- [13]. A.Wilbik, J.M.Keller, and J.C.Bezdek, "Linguistic prototypes for data from eldercare residents," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 1, pp. 110-123, 2013.
- [14]. T.Altıntop, R.R.Yager, D.Akay, F.E.Boran, and M.Ünal, "Fuzzy Linguistic Summarization with Genetic Algorithm: An Application with Operational and Financial Healthcare Data," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 25, no. 04, pp. 599-620, 2017.
- [15]. J.Kacprzyk, S.Zadrozny, A.Wilbik, "Using a genetic algorithm to derive a linguistic summary of trends in numerical time series," in *2006 International Symposium on Evolving Fuzzy Systems*, 2006, pp. 137-142: IEEE.
- [16]. C.H.Nguyen, W.Wechler, "Extended hedge algebras and their application to fuzzy logic," *Fuzzy sets and systems*, vol. 52, no. 3, pp. 259- 281, 1992.
- [17]. C.H.Nguyen and V.L.Nguyen, "Fuzziness measure on complete hedge algebras and quantifying semantics of terms in linear hedge algebras," *Fuzzy sets and Systems*, vol. 158, no. 4, pp. 452-471, 2007.
- [18]. C.H.Nguyen, T.S.Tran, and D.P.Pham, "Modeling of a semantics core of linguistic terms based on an extension of hedge algebra semantics and its application," *Knowledge-Based Systems*, vol. 67, pp. 244-262, 2014.
- [19]. C.H.Nguyen, V.T. Hoang, V.L.Nguyen, "A discussion on interpretability of linguistic rule based systems and its application to solve regression problems," *Knowledge-Based Systems*, vol. 88, pp. 107-133, 2015.
- [20]. C.H.Nguyen, N.L.Vu, X.V.Le, "Optimal hedge-algebras-based controller: Design and application," *Fuzzy Sets and Systems*, vol. 159, no. 8, pp. 968-989, 2008.



- [21]. H.L. Bui, C.H. Nguyen, N.-L. Vu, C.H. Nguyen, "General design method of hedge-algebras-based fuzzy controllers and an application for structural active control," *Applied Intelligence*, vol. 43, no. 2, pp. 251-275, 2015.
- [22]. C.H. Nguyen, W. Pedrycz, T.L. Duong, T. S. Tran, "A genetic design of linguistic terms for fuzzy rule based classifiers," *International Journal of Approximate Reasoning*, vol. 54, no. 1, pp. 1-21, 2013.
- [23]. N.D. Hieu, N.C. Ho, V.N. Lan, "Enrollment forecasting based on linguistic time series," *Journal of Computer Science Cybernetics*, vol. 36, no. 2, pp. 119-137, 2020.
- [24]. N.D. Hieu, N.C. Ho, and V.N. Lan, "An efficient fuzzy time series forecasting model based on quantifying semantics of words," in *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, 2020, pp. 1-6: IEEE.
- [25]. N.V. Tinh, N.C. Dieu, N.T. Duy, and T.T. Thanh, "Improved Fuzzy Time Series Forecasting Model Based on Optimal Lengths of Intervals Using Hedge Algebras and Particle Swarm Optimization," *Advances in Science, Technology and Engineering Systems Journal*, vol. 6, no. 1, pp. 1286-1297, 2021.
- [26]. B.K. Phan and V.T. Nguyen, "Hedge-Algebras-Based Controller for Mechanisms of Relative Manipulation," *International Journal of Precision Engineering Manufacturing*, vol. 19, no. 3, pp. 377-385, 2018.
- [27]. N.T. Duy and V.D. Vuong, "Designing hedge algebraic controller and optimizing by genetic algorithm for serial robots adhering trajectories," *Journal of Computer Science Cybernetics*, vol. 36, no. 3, pp. 265-283, 2020.
- [28]. C. H. Nguyen, W. Pedrycz, and V. N. Huynh, "A construction of sound semantic linguistic scales using 4-tuple representation of term semantics," *International Journal of Approximate Reasoning*, vol. 55, no. 3, pp. 763-786, 2014.
- [29]. C. H. Nguyen, N. L. Vu, and X. V. Le, "An interpolative reasoning method based on Hedge Algebras and its application to a problem of fuzzy control," in *Proceedings of the 10th WSEAS international conference on Computers*, 2006, pp. 13-15.
- [30]. N. D. Anh, H. L. Bui, N. L. Vu, and D. T. Tran, "Application of hedge algebra-based fuzzy controller to active control of a structure against earthquake," *Structural Control Health Monitoring*, vol. 20, no. 4, pp. 483- 495, 2013.



- [31]. H.L. Bui, D.T. Tran, and N. L.Vu, "Optimal fuzzy control of an inverted pendulum," *Journal of Vibration and Control Engineering Practice*, vol. 18, no. 14, pp. 2097-2110, 2012.
- [32]. Thi Lan Pham, Cam Ha Ho, and Cat Ho Nguyen, "Linguistic summarization based on the inherent semantics of linguistic words", *International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making*, 2018, pp. 15-26.