QUAN HỆ CÚ PHÁP VÀ NGỮ NGHĨA TRONG XÂY DỰNG PHƯƠNG PHÁP LẤY QUYẾT ĐỊNH ĐA TIÊU CHUẨN

Nguyễn Cát Hồ[1], Nguyễn Thanh Thủy[2], Nguyễn Văn Long[3], Lê Ngọc Hưng[4,*]

[1]Viện Công nghệ thông tin, Viện KHCNVN, 18 Hoàng Quốc Việt, Cầu Giấy, Hà Nội
[2]Trường Đại học Công nghệ, VNU
[3]Trường Đại học Giao thông Vận Tải, Hà Nội
[4]Trường Đại học Sài gòn, Tp Hồ Chí Minh

[*]Email: *lengochunsg291958@gmail.com*

Đến Toà soạn: 15/5/2011; Chấp nhận đăng: 16/4/2012

TÓM TẮT

Việc biểu diễn ngữ nghĩa các từ ngôn ngữ bằng chỉ số của chúng trong thang điểm ngôn ngữ trong một số cách tiếp cận hiện nay có nhược điểm là chúng là những giá trị không ổn định, phụ thuộc vào vị trí và số lượng từ sử dụng trong thang điểm. Bài báo đề xuất cơ sở nghiên cứu xây dựng thang điểm dựa trên việc thảo luận mối quan hệ cú pháp và ngữ nghĩa trong lô gic hình thức. Một số nguyên tắc được đề xuất cho việc xây dựng thang điểm phù hợp với bài toán ra quyết định theo nhóm chuyên gia, trong đó có nguyên tắc đối với toán tử kết nhập, và chỉ ra rằng biểu diễn bộ 4 ngữ ngĩa ngôn ngữ dựa trên ĐSGT đáp ứng được tốt các nguyên tắc đề ra. Một số ví dụ về bài toán quyết định sẽ minh chứng cho cách tiếp cận mới.

Từ khóa: thang điểm ngôn ngữ, lấy quyết định đa tiêu chuẩn, đại số gia tử, ngữ nghĩa khoảng, toán tử kết nhập

1. ĐẶT VẤN ĐỀ

Bài toán quyết định có vai trò rất quan trọng đối với việc quản lí ở mọi lĩnh vực. Đây là bài toán khó do nó luôn đặt trong bối cảnh phức tạp như môi trường cấu trúc yếu với thông tin không chính xác, không chắc chắn hay thông tin ngôn ngữ mờ. Việc lấy quyết định như vậy thường dựa vào ý kiến các chuyên gia. Điều này dẫn đến bài toán lấy quyết định mờ dựa vào nhóm chuyên gia. Vấn đề ra quyết định mờ, với ý nghĩa và tính chất như vậy, thu hút sự quan tâm mạnh mẽ của cộng đồng nghiên cứu và, vì vậy, có nhiều cách tiếp cận khác nhau.

Trong [1, 2, 3], và có thể tiện tham khảo trong [4], các tác giả đã nêu tổng quan một số cách tiếp cận. Cơ sở thúc đẩy việc nghiên cứu trong bài báo này xuất phát từ việc quan sát, phân tích một số vấn đề bất hợp lí căn bản trong cách tiếp cận sử dụng toán tử lấy trung bình ngôn ngữ có thứ tự và có trọng số LOWA (Linguistic Ordered Weighted Average) [6] và cách tiếp cận dựa trên biểu diễn 2-bộ (2-tuple representation). Từ nhu cầu thực tế, cả hai cách tiếp cận này đều giả thiết rằng thang điểm biểu thị ý kiến đánh giá của các chuyên gia theo một số tiêu chuẩn được biểu thị bằng các từ ngôn ngữ lấy trong tập $\boldsymbol{S} = \{s_0, \ldots, s_g\}$, được sắp tuyến tính theo ngữ nghĩa

của chúng. Quy ước rằng $s_i < s_j$, với $i < j$. Vì chưa có cơ sở toán học tính toán trực tiếp trên các từ nên người ta mượn cấu trúc tính toán của đoạn [0, g], bao hàm các chỉ số của chúng, để thực hiện các phép tính số học, tương tự như tính toán số học đối với thang điểm số.

Có thể nhận thấy việc dùng chỉ số để tính toán thay cho ngữ nghĩa của các từ ngôn ngữ trong thang điểm là khiên cưỡng và mất mát nhiều thông tin, vì theo cách biểu thị kí hiệu các từ được quy ước trong thang điểm, chỉ số của chúng chỉ phản ánh được thứ tự của ngữ nghĩa của các từ. Rõ ràng là ngữ nghĩa của các từ trong thang điểm ngôn ngữ phản ánh ngữ nghĩa nhiều hơn hẳn so với các chỉ số của chúng. Chẳng hạn các chỉ số phân bố đều trên thang điểm, trong khi thang điểm ngôn ngữ có phân bố không đều theo ngữ nghĩa của chúng. Ngữ nghĩa của từ "*xuất sắc*" có tính riêng biệt (specificity) nhiều hơn so với "*giỏi*" và do vậy khoảng các giá trị thực tương ứng với ngữ nghĩa của "*xuất sắc*" nhỏ hơn so với khoảng các giá trị tương ứng với "*giỏi*".

Trên quan điểm lô gic, việc làm rõ khái niệm cú pháp và ngữ nghĩa, cũng như mối quan hệ "triết lí" chặt chẽ giữa chúng là cực kì quan trọng. Ngữ nghĩa có tính trừu tượng, còn cú pháp đề cập đến kí hiệu và quy tắc xây dựng chúng. Tính toán trên các kí hiệu là một dãy liên tiếp các phép biến đổi trên các kí hiệu. Vì vậy có hai vấn đề đặt ra:

(i) Các kí hiệu cần biểu thị được đầy đủ ngữ nghĩa mà chúng mang hay được con người gán cho chúng trong một lĩnh vực ứng dụng đang được xem xét.

(ii) Phương pháp hay thuật toán tính toán liệu có cho kết quả cũng ở dạng kí hiệu có ngữ nghĩa phù hợp với mong muốn của người thiết kế phương pháp hay không.

Kí hiệu và ngữ nghĩa là hai mặt của một vấn đề có quan hệ khăng khít với nhau. Chẳng hạn như mối quan hệ hình - bóng giữa kí hiệu của các từ và các câu trong ngôn ngữ thường ngày của con người. Trong mối quan hệ này cú pháp chỉ là xâu các kí tự (các từ, câu), cùng với các quy tắc sinh các xâu, còn ngữ nghĩa của chúng được con người gán các sự vật, hiện tượng trong thế giới thực cho chúng. Hay, nói khác đi, ngữ nghĩa của một từ trong ngôn ngữ tự nhiên là một tập các sự vật, hiện tượng mà nó ám chỉ. Tuy nhiên, trong thực tiễn của môi trường thông tin mờ không phải lúc nào chúng ta cũng xác định được rõ ràng mối quan hệ này, do tập các sự vật, hiện tượng mà một từ ám chỉ khó có thể được xác định rõ ràng do chúng có danh giới không rõ ràng. Có vẻ như thực tế này được xem như là tự nhiên và khách quan và vì vậy nhiều khi vấn đề (i) nêu trên không được chú ý đúng mức cần thiết hay bị lãng quyên. Do vậy các cách tiếp cận dựa trên toán tử LOWA và phương pháp biểu diễn bộ 2 chỉ sử dụng chỉ số của các từ ngôn ngữ trong thang điểm để biểu thị ngữ nghĩa của các từ ngôn ngữ, những thông tin rất ít ỏi về ngữ nghĩa của chúng.

Những thiếu sót trên cũng là hệ quả của việc miền các giá trị ngôn ngữ chưa được hình thức hóa để mô hình hóa ngữ nghĩa định tính trong cách tiếp cận của lí thuyết tập mờ. Vì vậy có sự tách biệt giữa cú pháp và ngữ nghĩa của các giá trị ngôn ngữ trong lí thuyết tập mờ khi mà các đặc trưng bản chất này chưa được phát hiện. Mối liên hệ này chỉ được thực hiện thông qua cảm nhận trực quan của người nghiên cứu ứng dụng khi xây dựng hay gán các hàm thuộc cho các từ ngôn ngữ.

Những thiếu sót trên thúc đẩy việc thảo luận chi tiết vấn đề cú pháp và ngữ nghĩa của thang điểm ngôn ngữ trong việc nghiên cứu một số nguyên tắc trong xây dựng thang điểm và, đặc biệt, trong xây dựng các toán tử kết nhập (aggregation operator). Trên cơ sở đó chúng tôi chứng tỏ phương pháp biểu diễn ngữ nghĩa bộ 4 dựa trên cách tiếp cận ngữ nghĩa ngôn ngữ của đại số gia tử (ĐSGT) có nhiều ưu điểm để khắc phục các yếu điểm này.

Thang điểm ngôn ngữ ngữ nghĩa bộ 4 đã được nghiên cứu trong [4] nhưng chưa được thảo luận và khẳng định rõ ràng và trực diện trong tương quan với cú pháp. Ngoài ra trong bài báo đó

cũng chưa chứng tỏ được tính hơn hẳn của phương pháp biểu diễn 4-bộ thông qua việc nghiên cứu so sánh với các ví dụ có tính thuyết phục.

Phần còn lại của bài báo được trình bày trong 3 mục. Trong Mục 2, mối quan hệ cú pháp và ngữ nghĩa sẽ được thảo luận làm cơ sở đề xuất 3 nguyên tắc xây dựng thang điểm ngôn ngữ. Mục 3 dành cho việc chứng tỏ thang điểm ngôn ngữ với biểu diễn ngữ nghĩa bộ 4 có thể đáp ứng được 3 nguyên tắc đề ra, trong khi các cách tiếp cận khác có những yếu điểm. Trong Mục 4 việc so sánh giữa các phương pháp tiếp cận sẽ được nghiên cứu để chỉ ra sự khác biệt trên quan điểm kết quả của phương pháp thỏa mãn 3 nguyên tắc đề ra tốt hơn, đáng tin cậy hơn. Mục 5 dành cho kết luận.

2. MỘT SỐ NGUYÊN TẮC XÂY DỰNG THANG ĐIỂM

Như đã đề cập trong Mục 1, có những vấn đề ngữ nghĩa và phương pháp biểu diễn nổi lên cần giải quyết khi xây dựng thang điểm ngôn ngữ mờ. Trong mục này chúng ta sẽ thảo luận một số nguyên tắc đảm bảo tính chính đáng, tính hợp lí của việc xây dựng chúng. Những yếu điểm của một số cách tiếp cận hiện tại trong xây dựng thang điểm ngôn ngữ trên các tạp chí quốc tế, chẳng hạn trong [1, 2, 3], cho thấy việc thảo luận này trở nên cần thiết và quan trọng.

2.1. Phân tích cú pháp và ngữ nghĩa của một số loại thang điểm

Cũng như đối với vấn đề về tính đúng đắn của lập luận trong lô gic hay của chương trình máy tính, cú pháp và ngữ nghĩa đóng vai trò nền tảng. Cú pháp đề cập đến các xâu kí hiệu, trong lôgic gọi là các công thức, và các quy tắc xây dựng các công thức đúng (well-formed-formulas), còn ngữ nghĩa là các sự vật, hiện tượng trong thế giới thực được gán cho chúng. Xâu kí hiệu là vật - mang mang ngữ nghĩa, thông tin và là phương tiện con người hay thiết bị, máy tính dùng để truyền tin, xử lí thông tin. Ngữ nghĩa mà một xâu kí hiệu mang ám chỉ các sự vật hiện tượng mà con người gán cho nó, và được gọi là các cái bị trỏ trong thế giới thực của kí hiệu. Ví dụ, từ "sông" là kí hiệu tiếng Việt mà ngữ nghĩa của nó được hình thành qua các cái bị trỏ gồm các con sông trong thế giới thực được cộng đồng người Việt gán cho nó.

Mối quan hệ giữa cú pháp và ngữ nghĩa trong ngôn ngữ của một cộng đồng được hình thành cùng với lịch sử tồn tại của họ trở nên hiển nhiên đến mức đối với hầu hết mọi người không cảm nhận thấy sự tồn tại mối quan hệ này. Điều này có thể gây ra hệ lụy. Chẳng hạn một số nghịch lí được phát hiện trong lí thuyết tập hợp của Cantor (được gọi là lí thuyết tập hợp "ngây thơ") chỉ ra rằng sự không tường minh về ngữ nghĩa của một số khái niệm trong lí thuyết này chính là thủ phạm sinh ra những nghịch lí đó. Có lẽ cũng vì vậy mà tồn tại một khoảng cách tách biệt đáng kinh ngạc giữa kí hiệu các từ ngôn ngữ và ngữ nghĩa chúng cần mang trong xây dựng một số thang điểm ngôn ngữ để giải quyết bài toán lấy quyết định đa tiêu chuẩn, như chúng ta sẽ phân tích dưới đây. Vì vậy, việc nghiên cứu nhằm góp phần làm sáng tỏ mối quan hệ thích đáng giữa cú pháp và ngữ nghĩa trong các bài toán ra quyết định mờ có ý nghĩa quan trọng. Điều này cũng đóng vai trò rất quan trọng trong lĩnh vực nghiên cứu trí tuệ nhân tạo. Giới hạn trong vấn đề xây dựng thang điểm ngôn ngữ và các phương pháp ra quyết định, hai khía cạnh của mối quan hệ này cần tách biệt:

a) Ngữ nghĩa của các từ, tức các xâu kí hiệu, trong thang điểm (kí hiệu số, kí hiệu giá trị ngôn ngữ mờ, …) mà nó biểu thị mối quan hệ định lượng hoặc định tính giữa các giá trị của thang điểm, dựa vào đó chúng mang những thông tin cho phép so sánh giá trị này

hơn giá trị kia một lượng định tính hay định lượng nhất định. Ví dụ, một học sinh được đánh giá *xuất sắc* sẽ hơn một học sinh được đánh giá là *khá* một lượng định tính là bao nhiêu là thích đáng trong một ứng dụng.

b) Ngữ nghĩa mà các chuyên gia trong hội đồng đánh giá các phương án lựa chọn (đối tượng cần đánh giá) gán cho một từ nào đó trong thang điểm, dựa vào ngữ nghĩa của từ đó trong thang điểm (ngữ nghĩa theo điểm a)), để biểu thị mức độ thỏa mãn của mỗi đối tượng đối với một tiêu chuẩn đánh giá được xét.

Ngữ nghĩa theo điểm b) là một vấn đề phức tạp, phi cấu trúc và phụ thuộc vào từng ứng dụng cụ thể (và vì vậy cần sử dụng phương pháp chuyên gia). Do vậy trong nghiên cứu này chúng ta chỉ quan tâm nghiên cứu mối quan hệ cú pháp và nghữ nghĩa của thang điểm theo điểm

1) Cú pháp và ngữ nghĩa trong thang điểm ngôn ngữ

Thang điểm ngôn ngữ là một tập hữu hạn gồm một số lẻ các từ ngôn ngữ mờ có thứ tự tuyến tính [2]. Số lẻ các từ là cần thiết trong một số cách tiếp cận để có từ biểu thị ý kiến đánh giá giữa *khá* và *kém*. Một ví dụ thang điểm ngôn ngữ là tập {*kém*, *trung_bình*, *khá*, *giỏi*, *xuất_sắc*} với quy ước các chuyên gia sẽ dùng các từ ngôn ngữ trong thang điểm để đánh giá thay vì dùng số thực. Sự khác biệt quan trọng của thang điểm này so với thang điểm số là thang điểm số gồm các giá trị có ngữ nghĩa chính xác, trong khi các giá trị ngôn ngữ lại là mờ, không chính xác và chúng không có cấu trúc cho phép tính toán.

Có những lí do đòi hỏi việc xây dựng thang điểm ngôn ngữ:

- Việc đánh giá cho điểm của các chuyên gia thường trong môi trường thông tin và dữ kiện mờ, khó đưa ra những định giá, chính kiến bằng giá trị chính xác do sự tương phản của hai môi trường thông tin.

- Con người nói chung thường xuyên và quen sử dụng, xử lí các thông tin ngôn ngữ mờ. Vì vậy, các chuyên gia dễ đưa ra chính kiến của mình biểu thị bằng từ ngôn ngữ mờ. Những nghiên cứu thực nghiệm chỉ ra rằng sử dụng thông tin ngôn ngữ mờ có tỷ lệ cho chính kiến thống nhất cao hơn so với việc sử dụng thông tin số (anh ta cao 1,65 m), khoảng số (anh ta cao khoảng từ 1,62 m đến 1,67 m).

Với lí do trên, mặc dù việc tính toán để kết nhập ý kiến đánh giá của các chuyên gia trên thang điểm ngôn ngữ gặp nhiều thách thức, lĩnh vực này vẫn thu hút được nhiều chú ý của nhiều nhà nghiên cứu với định hướng ứng dụng thiết thực [3, 8, 9]. Nhiều cách tiếp cận đã được đưa ra để vừa đảm bảo biểu thị ngữ nghĩa của các từ ngôn ngữ trong thang điểm, vừa cho phép định nghĩa các phép kết nhập các ý kiến đánh giá theo nhiều tiêu chuẩn khác nhau hoặc/và của một nhóm các chuyên gia. Đáng tiếc là một số hướng tiếp cận đã không làm rõ hay nêu đích danh vấn đề biểu diễn ngữ nghĩa của các từ ngôn ngữ. Vì vậy, chúng ta chỉ có thể phân tích bản chất ngữ nghĩa của các từ ngôn ngữ bằng việc xem xét bản chất của phương pháp tính toán lựa chọn quyết định dựa trên đại lượng nào trong từng cách tiếp cận.

2) Cách tiếp cận tập mờ

Như đã trình bày ở phần mở đầu, chúng ta tính toán trên các xâu kí hiệu, nhưng tập các từ ngôn ngữ lại không có cấu trúc tính toán để kết nhập các ý kiến chuyên gia. Vì vậy, trong cách tiếp cận mờ, các kí hiệu ngôn ngữ được biểu thị bằng các tập mờ mà hàm thuộc của chúng là một đối tượng tính toán được. Nghĩa là ngữ nghĩa của mỗi từ ngôn ngữ *được biểu thị bằng các giá trị số trong miền tham chiếu được xem là phù hợp với ngữ nghĩa của từ ngôn ngữ với một độ*

phù hợp bằng giá trị hàm thuộc ở các giá trị số tương ứng. Như vậy quan hệ cú pháp và ngữ nghĩa của các từ trong thang điểm được xác lập bằng việc xây dựng các hàm thuộc gán cho nó.

Tuy nhiên, cũng như bất kì thang điểm nào khác, một yêu cầu cốt yếu là phải định lượng được "khoảng cách" giữa các từ trong thang điểm ngôn ngữ để phản ánh sự gần hay xa nhau giữa ngữ nghĩa định tính của các từ ngôn ngữ. Việc định nghĩa khoảng cách giữa các tập mờ là một bài toán khó và có rất nhiều giải pháp khác nhau. Ngoài ra việc kết nhập các ý kiến đánh giá biểu thị bằng các từ trong thang điểm được thực hiện trên các hàm thuộc gán cho chúng nhờ áp dụng *nguyên lí thác triển* (*extension principle*) đối với phép kết nhập số học được lựa chọn. Tuy nhiên, kết quả thu được là một tập mờ cần phải được chuyển về một từ trong thang điểm để biểu thị kết quả của kết nhập. Việc này được thực hiện bằng các phương pháp xấp xỉ, gọi là các phương pháp xấp xỉ ngôn ngữ (*linguistic approximation*), vì bản chất của phương pháp là tìm một cách đánh giá để xác định tập mờ của từ ngôn ngữ nào trong thang điểm *gần* với tập mờ kết quả nhất.

Ngoài ra, ngữ nghĩa tập mờ của các từ cũng hàm chứa những nhược điểm căn bản:

(i) Ngữ nghĩa tập làm lu mờ thứ tự tự nhiên giữa các từ trong thang điểm ngôn ngữ, do các tập mờ của chúng đòi hỏi được xây dựng trùng dẫm lên nhau. Việc làm lu mờ thứ tự của các từ ngôn ngữ có thể làm sai lệch kết quả của các phép toán kết nhập.

(ii) Ngữ nghĩa biểu thị qua hàm thuộc được định nghĩa qua từng điểm của miền tham chiếu dẫn đến mọi tính toán đều phải tính theo từng điểm. Vậy, phép kết nhập cũng buộc phải tính toán trên từng điểm dựa trên nguyên lí thác triển. Khó có cảm nhận trực giác về kết quả cụ thể của phép kết nhập, vì nó phụ thuộc vào nhiều yếu tố phức tạp:

- Việc xây dựng tập mờ cho thang điểm: *bài toán hàm thuộc*.
- Vì kết quả của phép kết nhập là một tập mờ thường khác với các tập mờ của các từ nên đòi hỏi phải giải *bài toán xấp xỉ ngôn ngữ.*

3) Cách tiếp cận tính toán từ (*word computing*) *– Toán tử LOWA*

Để tránh cách tiếp cận phức tạp trên, tác giả [2] đã đưa ra cách tiệp cận tính toán "trực tiếp" trên các từ ngôn ngữ bằng toán tử LOWA (Linguistic Odered Weighted Averaging). Ta sẽ phân tích xem việc tính toán phép kết nhập trong này có thật trực tiếp trên các từ hay không. Muốn vậy, ta hãy phân tích thực chất định nghĩa của toán tử LOWA, để qua đó thấy bản chất ngữ nghĩa của các từ trong thang điểm đối với cách tiếp cận này.

Giả sử $\{a_1, \ldots, a_m\}$ là tập gồm m từ ngôn ngữ không nhất thiết khác nhau của thang điểm ngôn ngữ $\boldsymbol{S} = \{s_0, \ldots, s_g\}$ để thực hiện việc kết nhập để tìm kết luận cho việc lấy quyết định. Khi đó giá trị của toán tử LOWA $\phi(a_1, \ldots, a_m)$ được tính bằng tổ hợp lồi C như sau [2]:

$$\phi(a_1, \ldots, a_m) = W \odot B^T = C^m\{w_k, b_k; k = 1 \,..\, m\} = w_1 \odot b_1 \oplus (1-w_1) \odot C^{m-1}\{w_{2,k}, b_k; k = 2 \,..\, m\}$$

trong đó,

+ $W = (w_1, \ldots, w_m)$, $B = (b_1, \ldots, b_m) = (a_{\sigma(1)}, \ldots, a_{\sigma(m)})$, với σ là một hoán vị của các phần tử 1, …, m sao cho $a_{\sigma(i)} \geq a_{\sigma(j)}$ với mọi $j \geq i$;

+ $w_k \in [0,1]$, $\sum_{k=1,\ldots,m} w_k = 1$ và $w_{2,k} = w_k / \sum_{j=2,\ldots,m} w_j = w_k/(1 - w_1)$.

+ Các phép $\oplus$ và $\odot$ là phép toán hình thức trên các từ ngôn ngữ không có cấu trúc tính toán. Tiếc rằng ý nghĩa của chúng chưa được chỉ ra rõ trong [2, 5] mà chúng ta chỉ có thể hiểu nó qua các tính toán cụ thể dưới đây.

Bằng quy nạp ta thu được công thức sau:

$$\phi(a_1, \ldots, a_m) = W \odot B^T = C^m\{w_k, b_k ; k = 1, \ldots, m\} = w_1 \odot b_1 \oplus w_2 \odot b_2 \oplus \ldots \oplus w_m \odot b_m$$

với b_k là các kí hiệu các từ ngôn ngữ vốn chúng ta không thể tính toán số học đối với chúng, nên ngữ nghĩa của các phép tính $\oplus$ và $\odot$ được hiểu như sau.

Tham khảo [4], với $m = 2$, tổ hợp lồi C^2 với $b_1 = s_j$ và $b_2 = s_i$ được tính như sau:

$$C^2 = w_1 \odot s_j \oplus (1 - w_1) \odot s_i = s_h \in \boldsymbol{S},$$

với $h = i + \text{round}(w_1(j - i))$. Trường hợp tổng quát ta có:

$$\phi(a_1, \ldots, a_m) = W \odot B^T = w_1 \odot s_{j(1)} \oplus w_2 \odot s_{j(2)} \oplus \ldots \oplus w_m \odot s_{j(m)} = s_h,$$

với $h = \text{round}[w_1 . j(1) + w_2 . j(2) + \ldots + w_m . j(m)] \in [0, g]$.

Như vậy, thực chất ý nghĩa các phép tính $\oplus$ và $\odot$ phải được hiểu chính là *phép nhân số học* và *phép cộng số học* đối với các *chỉ số* của từ ngôn ngữ và các trọng số, sau đó kết quả được làm tròn để tìm từ ngôn ngữ trong thang điểm làm kết quả của phép LOWA. Cho nên ta có thể kết luận rằng *ngữ nghĩa của các từ ngôn ngữ có thứ tự trong cách tiếp cận này được biểu thị bằng chỉ số của chúng trong dãy được sắp thứ thụ theo ngữ nghĩa*.

Có thể nhận thấy mối quan hệ cú pháp và ngữ nghĩa của các từ như vậy rất nghèo nàn, vì các chỉ số chỉ chỉ ra được mối quan hệ thứ tự của các từ trong thang điểm. Ngữ nghĩa của các từ trong thang điểm phải có những đặc trưng khác phong phú hơn nhiều đặc trưng thứ tự. Chẳng hạn, các giá trị chỉ số phân bố đều nhưng giá trị định tính của các từ, ví dụ như trong đánh giá học sinh, lại cho chúng ta hiểu "khoảng cách" giữa các từ không phân bố đều.

Ngoài ra, ngữ nghĩa của từ, tuy không chính xác, nhưng được xác định định tính, không phụ thuộc các từ khác hiện diện hay vắng mặt trong thang điểm, trong khi ngữ nghĩa định lượng biểu thị qua giá trị các chỉ số của chúng lại phụ thuộc vào số lượng các từ trong thang điểm.

4) Cách tiếp cận tính toán từ với biểu diễn bộ 2 (2-tuple representation)

Cách tiếp cận 3) ở trên có yếu điểm là việc làm tròn kết quả tính toán để xác định chỉ số của từ ngôn ngữ kết quả kết nhập sẽ làm mất mát thông tin vì nó không ghi nhận được kết quả lớn hơn hay nhỏ hơn số nguyên được làm tròn. Trong trường hợp hai kết quả tính toán kết nhập cho hai phương án lựa chọn với một nhỏ hơn và một lớn hơn, nhưng đều được làm tròn về chỉ số của nhãn ngôn ngữ s_j, sẽ làm mất thông tin cho ta biết rằng có một phương án lựa chọn được đánh giá tốt hơn phương án còn lại. Vì vậy, người ta đưa ra biểu diễn từ ngôn ngữ bằng bộ 2 [2, 3], tức là một cặp (s_j, r_j) gồm một từ ngôn ngữ và một giá trị số $r_j \in [-0,5; 0,5]$ xác định bằng đẳng thức $r_j = \sigma - j$, trong đó σ là kết quả của phép tính kết nhập số học trên các chỉ số của các từ ngôn ngữ. Trong [2] giá trị r_j được gọi là một sự tịnh tiến kí hiệu (symbolic translation). Khái niệm này thể hiện rõ nhất khi ứng dụng vào bài toán phân lớp mờ: Tập mờ tam giác ứng với từ s_j có đỉnh tại j, sẽ được tịnh tiến đến giá trị $(j + r_j)$.

Như vậy, ngoài việc xét thêm thành phần r_j liên quan đến kết quả của phép kết nhập, yếu điểm về mối quan hệ giữa cú pháp và ngữ nghĩa trong cách tiếp cận này hoàn toàn giống như cách tiếp cận trên. Ngoài ra, khi đưa ra biểu diễn bộ 2 thì lại xuất hiện vấn đề về mối quan hệ giữa kí hiệu s_j và ngữ nghĩa của nó biểu thị bằng bộ 2 (s_j, r_j).

Tóm lại, giá trị số r_j được đưa vào chỉ để giải quyết việc kết quả kết nhập không trùng với giá trị chỉ số của các từ trong thang điểm. Triết lí cách tiếp cận này không thay đổi so với cách tiếp cận ở Điểm 3) và nó vẫn giữ nguyên những nhược điểm của cách tiếp cận ấy.

2.2. Một số nguyên tắc xây dựng thang điểm ngôn ngữ

Lí do chính thúc đẩy việc nghiên cứu này là mối quan hệ giữa cú pháp và ngữ nghĩa của các kí hiệu trong thang điểm ngôn ngữ không được quan tâm đến mức bị lãng quên như trong các cách tiếp cận đã đề cập ở điểm 3) và 4), Mục 2.1, vì chúng ta không thể lí giải được vì sao lại mượn chỉ số của các từ trong thang điểm thay cho ngữ nghĩa của chúng, ngoài lí do thang điểm ngôn ngữ không có cấu trúc tính toán. Quan điểm được đưa ra trong nghiên cứu này là phải tính toán trên biểu diễn ngữ nghĩa của các từ ngôn ngữ. Vì vậy, nếu các chỉ số của chúng không được xem là biểu diễn ngữ nghĩa chính đáng của các từ thì các cách tiếp cận 3) và 4) trở nên không thỏa đáng. Để bảo đảm tính thích đáng của các thang điểm ngôn ngữ, bài báo đưa ra một số nguyên tắc trong xây dựng thang điểm ngôn ngữ cho các bài toán ra quyết định, trên cơ sở nhấn mạnh vai trò quan trọng của ngữ nghĩa các từ.

*Nguyên tắc 1***:** *Bảo đảm ngữ nghĩa định lượng của các kí hiệu ngôn ngữ trong thang điểm phải phù hợp với ngữ nghĩa định tính của các từ ngôn ngữ trong ngôn ngữ tự nhiên.*

Trong môi trường thông tin mờ, việc bảo đảm thực hiện nguyên tắc này là một bài toán khó vì thông tin không chính xác, không chắc chắn và do đó không thể dễ dàng đưa ra các tiêu chuẩn rõ ràng, cụ thể trong việc xây dựng thang điểm ngôn ngữ, cùng với việc đưa ra phép kết nhập các giá trị trong thang điểm dựa trên nguyên tắc này. Việc không tuân thủ nguyên tắc này sẽ dẫn đến những thiếu sót ngoài ý muốn như đã phân tích đối với một số cách tiếp cận hiện có. Dù không có các tiêu chí cụ thể, nguyên tắc này vẫn có giá trị chỉ dẫn quan trọng trong việc thể hiện *mối liên hệ thích đáng* giữa ngữ nghĩa định tính và định lượng của các từ ngôn ngữ.

Hiện nay chúng ta chỉ có thể thực hiện việc kết nhập các từ ngôn ngữ khi biểu diễn được ngữ nghĩa định lượng của chúng. Nếu không chỉ ra được ngữ nghĩa định lượng biểu thị *thích đáng* ngữ nghĩa định tính của các từ ngôn ngữ, có thể có những mất mát thông tin trong phương pháp luận. Chẳng hạn, việc biểu thị ngữ nghĩa bằng các chỉ số của chúng trong một thang điểm ngôn ngữ có thứ tự sẽ mất mát nhiều thông tin, có thể làm cho kết quả lấy quyết định sai lệch như chúng ta sẽ chỉ ra ở Mục 4. Nó không có liên hệ *đủ thích đáng* về mặt ngữ nghĩa. Cho nên, khi bổ sung thêm các từ ngôn ngữ vào thang điểm, ngữ nghĩa của chúng thay đổi, trong khi ngữ nghĩa của từ ngôn ngữ là xác định. Nếu giả sử ta biện minh rằng các chỉ số của chúng không phải được sử dụng để biểu thị ngữ nghĩa của chúng thì chính là đã phạm phải Nguyên tắc 1 về căn bản: không được tính toán trên các giá trị định lượng mà không biểu thị ngữ nghĩa của các từ trong thang điểm.

Nguyên tắc 2: Biểu diễn định lượng ngữ nghĩa của các từ trong thang điểm phải phản ánh được những tính chất cốt yếu của ngữ nghĩa định tính của các từ ngôn ngữ.

Yêu cầu này là một đòi hỏi tự nhiên nhưng hiện tại, cũng như nhiều vấn đề trong nghiên cứu thông tin mờ khác, khó có thể đưa ra các tiêu chuẩn cụ thể để thể hiện như thế nào là "phản ánh được những tính chất cốt yếu của ngữ nghĩa định tính" của các từ ngôn ngữ. Tuy nhiên, nguyên tắc này đòi hỏi phải chỉ ra những đặc trưng định lượng có liên hệ thích đáng với ngữ

nghĩa định tính. Với đòi hỏi đó, chắc chắn việc lấy chỉ số để biểu thị ngữ nghĩa định lượng của các từ trong thang điểm là không thích đáng. Trong các công trình [2, 3] sử dụng cách tiếp cận 2 và 3, các tác giả đã không đề cập và chỉ ra mối liên hệ giữa các chỉ số của các từ trong thang điểm với ngữ nghĩa của chúng.

Có 2 cách tiếp cận có thể được xem là biểu diễn ngữ nghĩa định lượng phản ánh được một số tính chất cốt yếu ngữ nghĩa định tính của các từ ngôn ngữ:

(i) Các cách tiếp cận với ngữ nghĩa các từ ngôn ngữ biểu diễn bằng tập mờ [1]. Trong các lĩnh vực ứng dụng khác nhau, lí thuyết tập mờ đã đem lại những thành tựu rất to lớn.

(ii) Cách tiếp cận với ngữ nghĩa định lượng của các từ được biểu thị dựa trên việc định lượng ĐSGT [4, 7] mà sẽ được thảo luận chi tiết trong bài báo này.

Nguyên tắc 3: *Một cơ sở hình thức hóa đúng đắn cho các phép tính kết nhập* (*aggregation operators*) *thực hiện trên các giá trị của thang điểm, và nó phải đóng đối với kết quả kết nhập, tức là kết quả của phép kết nhập cũng là phần tử của thang điểm ngôn ngữ. Ngoài ra cần đòi hỏi rằng phép kết nhập các giá trị của thang điểm đơn điệu tăng thực sự, tức là:*

$$\boldsymbol{x} \le \boldsymbol{y} \Rightarrow Agg(\boldsymbol{x}) \le Agg(\boldsymbol{y}) \text{ và } \boldsymbol{x} \ne \boldsymbol{y} \Rightarrow Agg(\boldsymbol{x}) \ne Agg(\boldsymbol{y})$$

trong đó $\boldsymbol{x}$ *và* $\boldsymbol{y}$ *là các vec-tơ của các giá trị trong thang điểm của các tiêu chuẩn ra quyết định.*

Trong việc phát triển phương pháp lấy quyết định mờ theo nhóm chuyên gia, một biểu diễn ngữ nghĩa định lượng của các từ ngôn ngữ là chấp nhận được nếu dựa vào đó ta có thể xây dựng các phép tính kết nhập hợp lí để tổng hợp các ý kiến chuyên gia. R. Yager đã có một loạt bài nghiên cứu sâu sắc về các phép tính kết nhập [17, 18]. Nhìn chung, các phép kết nhập đều dựa trên họ các phép kết nhập này.

Tuy nhiên, nguyên tắc này được đưa ra còn để khẳng định khả năng bỏ ngỏ cho những cố gắng có thể có trong việc xây dựng các phép kết nhập trực tiếp các phần tử ngôn ngữ.

Sau đây chúng ta sẽ chỉ ra một thang điểm ngôn ngữ có thể được xem là được xây dựng tuân thủ được 3 nguyên tắc trên.

3. BIỂU DIỄN NGỮ NGHĨA BỘ 4 CỦA THANG ĐIỂM NGÔN NGỮ

Thang điểm bộ 3 đã được đề xuất và nghiên cứu trong [7] và thang điểm bộ 4 được phát triển trong [4], một sự mở rộng. Trong mục này chúng tôi sẽ phân tích mối quan hệ giữa cú pháp và ngữ nghĩa của thang điểm ngôn ngữ với biểu diễn ngữ nghĩa bộ 4 dựa trên ĐSGT theo 3 nguyên tắc đề xuất ở trên và chỉ ra rằng chỉ có ĐSGT thể hiện tốt mối quan hệ thực chất giữa cú pháp và ngữ nghĩa của các từ ngôn ngữ trong thang điểm.

3.1. Quan hệ cú pháp và ngữ nghĩa của thang điểm ngôn ngữ với biểu diễn bộ 4

3.1.1. Ngữ nghĩa dựa trên thứ tự của các từ trong một miền ngôn ngữ

Trong logic hình thức, kí hiệu ngôn ngữ tự nhiên là phương tiện được truyền và trao đổi qua các tín hiệu người với người, vốn nó không mang ngữ nghĩa. Ngữ nghĩa là các sự vật, sự kiện và hiện tượng được gán cho các kí hiệu, được gọi là các cái bị trỏ. Kí hiệu "con sông lớn" về cú pháp là một xâu các chữ cái, không mang nghĩa. Nó mang nghĩa thông qua những dòng sông ở thế giới thực được một cộng đồng người Việt gán cho nó. Ở mức độ trừu tượng hơn, "màu vàng" mang nghĩa thông qua những vật có màu vàng được gán cho nó.

Ngữ nghĩa các từ ngôn ngữ theo nghĩa này vẫn nằm ngoài khả năng hình thức hóa toán học để mô hình hóa và tạo dựng công cụ thao tác trên mô hình. Tuy nhiên, vẫn có những đặc trưng ngữ nghĩa định tính các từ ngôn ngữ cho phép hình thức hóa toán học.

Như trong Mục 1 đã đề cập, trong hoạt động quyết định của con người, quan trọng là chúng ta đánh giá được lựa chọn này tốt hơn lựa chọn kia, nghĩa là ngữ nghĩa dựa trên thứ tự là đủ cho việc lấy quyết định. Đòi hỏi như vậy trong lấy quyết định dẫn đến các từ ngôn ngữ dùng để mô tả tính chất của sự vật, hiện tượng có khả năng so sánh được với nhau. Ngôn ngữ của con người có đủ sắc thái phong phú, và cần phải đủ phong phú như vậy để mô tả sự vật hiện tượng muôn hình muôn vẻ trong thế giới thực để phục vụ việc lấy quyết định. Các từ nhấn gia tử được sinh ra để làm thay đổi các sắc thái ngôn ngữ, và như vậy nó thay đổi quan hệ thứ tự của từ trước và sau khi bị một gia tử tác động, cũng để phục vụ việc lấy quyết định.

Giới hạn việc nghiên cứu ngữ nghĩa của các từ trong việc biểu thị quan hệ thứ tự, ngữ nghĩa của một từ ngôn ngữ có thể được xác định qua các "sự kiện" về quan hệ thứ tự của nó với các từ còn lại trong một miền ngôn ngữ của thuộc tính, miền ngôn ngữ trở thành một ĐSGT, hay ĐSGT mô hình hóa ngữ nghĩa ngôn ngữ dựa trên quan hệ thứ tự. Như vậy, ĐSGT có thể được xem là mô hình hóa ngữ nghĩa định tính của các từ ngôn ngữ, hay *ngữ nghĩa định tính của các từ ngôn ngữ được hình thức hóa toán học bởi ĐSGT*. Do vậy, biểu diễn ngữ nghĩa định tính các từ dựa trên ĐSGT sẽ có cơ sở ngữ nghĩa định tính khách quan.

3.2.2. Biểu diễn bộ bốn và Nguyên tắc 2 đối với việc xây dựng thang điểm các từ ngôn ngữ

Xét ĐSGT tuyến tính $\mathbf{AX} = (\boldsymbol{X}, \boldsymbol{G}, \boldsymbol{C}, \boldsymbol{H}, \leq)$ của một biến ngôn ngữ $\mathbf{X}$ với $\boldsymbol{H}^+ = \{h_1, ..., h_p\}$ và $\boldsymbol{H}^- = \{h_{-1}, ..., h_{-q}\}$, thỏa mãn $h_1 < ... < h_p$ và $h_{-1} < ... < h_{-q}$, $p, q \geq 2$. Giả định bạn đọc ít nhất đã có những kiến thức tổng quan về ĐSGT, nếu không có thể xem tài liệu [6].

Giả sử thang điểm ngôn ngữ $\boldsymbol{S} = \{s_0, \ldots, s_g\}$ gồm các từ của ĐSGT $\mathbf{AX}$, $\{s_0, \ldots, s_g\} \subseteq \boldsymbol{X}$. Cũng như các cách tiếp cận đã phân tích trong Mục 2, các từ trong thang điểm ngôn ngữ cần được xây dựng biểu diễn ngữ nghĩa định lượng dựa trên việc lượng hóa ĐSGT [4, 12, 13, 14] và cho phép thực hiện được các phép kết nhập dễ dàng.

Bốn đặc trưng quan trọng sau đây bảo đảm biểu diễn ngữ nghĩa các từ được phong phú:

(i) Vì tập nền $\boldsymbol{X}$ của ĐSGT $\mathbf{AX}$ bảo toàn thứ tự ngữ nghĩa của các từ ngôn ngữ của biến ngôn ngữ $\mathbf{X}$, tập $\boldsymbol{H}(x)$ gồm những phần tử sinh ra từ x và các gia tử trong $\boldsymbol{H}$ sẽ xác định một khoảng của tập sắp thứ tự tuyến tính $(\boldsymbol{X}, \leq)$ chứa $\boldsymbol{H}(x)$ mà không chứa phần tử nào khác. Vì vậy, nó xác định một lân cận gồm những phần tử gần gũi ngữ nghĩa với x. Điều này phù hợp với ý nghĩa và chức năng của các gia tử khi sinh tập $\boldsymbol{H}(x)$. Với tính chất đó của tập $\boldsymbol{H}(x)$, nó được lấy làm mô hình về *tính mờ* của từ x và kích cỡ của nó được xem là *độ đo tính mờ* của x, kí hiệu là $fm(x) \in [0,1]$.

Họ các tập $\boldsymbol{H}(x)$ như vậy có cấu trúc và được lấy làm cơ sở cho hệ tiên đề xác định $fm(x)$:

(*fm1*) $fm(\boldsymbol{0}) = fm(\boldsymbol{W}) = fm(\boldsymbol{1}) = 0, fm(c^-) + fm(c^+) = 1$ và $\sum_{h \in H} fm(hx) = fm(x)$, $x \in \boldsymbol{X}$;

(*fm2*) $fm(hx)/fm(x)$ không phụ thuộc vào x gọi là độ đo tính mờ của h, kí hiệu là $\mu(h)$. Do đó, $fm(y) = \mu(h_m) \ldots \mu(h_1) fm(c)$, với $y = h_m \ldots h_1 c$ là biểu diễn chuẩn tắc của y.

(*fm3*) Đặt $\sum_{-q \leq i \leq -1} \mu(h_i) = \alpha$ và $\sum_{1 \leq i \leq p} \mu(h_i) = \beta$ ta có $\alpha + \beta = 1$ và $\alpha, \beta > 0$.

Ta thấy $fm(x)$ hoàn toàn xác định nếu chúng ta cho biết các giá trị $fm(c^-)$, $fm(c^+)$ và $\mu(h)$, $h \in \boldsymbol{H}(x)$, được gọi là *các tham số tính mờ* của **X**. Các tham số này rất quan trọng cho tính toán các đặc trưng định lượng khác.

(ii) Khoảng tính mờ $\mathfrak{I}(x)$ của $x \in \boldsymbol{X}$: mỗi $x \in \boldsymbol{X}$ được gán một khoảng $\mathfrak{I}(x) \subseteq [0,1]$ được xác định dựa trên tập $\boldsymbol{H}(x)$ sao cho độ dài khoảng $|\mathfrak{I}(x)| = fm(x)$ và các khoảng tính mờ của các từ có cùng độ dài, được tính bằng số các lần xuất hiện các gia tử trong biểu diễn chuẩn tắc của chúng, được sắp xếp theo cùng thứ tự giữa chúng: Các khoảng trong dãy sau là phân hoạch của $\mathfrak{I}(x)$ và tuân theo thứ tự như sau:

$$\mathfrak{I}(h_{-q}x) \le \mathfrak{I}(h_{-q+1}x) \le ... \le \mathfrak{I}(h_{-1}x) \le \mathfrak{I}(h_1x) \le \mathfrak{I}(h_2x) \le ... \le \mathfrak{I}(h_px), \text{ nếu } Sign(h_px) = +1$$

và $\mathfrak{I}(h_px) \le \mathfrak{I}(h_{p-1}x) \le ... \le \mathfrak{I}(h_1x) \le \mathfrak{I}(h_{-1}x) \le \mathfrak{I}(h_{-2}x) \le ... \le \mathfrak{I}(h_{-q}x)$, nếu $Sign(h_px) = -1$.

Đặt $\boldsymbol{X}_k = \{x \in \boldsymbol{X}: |x| = k\}$, $|x|$ là độ dài của từ x. Khi đó, họ các khoảng tính mờ $\{\mathfrak{I}(x): \in \boldsymbol{X}_k\}$ lập thành một phân hoạch của [0,1] và $\forall x, y \in \boldsymbol{X}_k$, $x \le y \Rightarrow \mathfrak{I}(x) \le \mathfrak{I}(y)$.

(iii) Giá trị định lượng của các từ: chúng được xác định bởi ánh xạ định lượng $\upsilon : \boldsymbol{X} \to [0,1]$. $\upsilon(x)$ được đặc trưng bởi hai tính chất:

- $\upsilon(x)$ là ánh xạ 1-1, bảo toàn thứ tự của tập $\boldsymbol{X}$ và tập ảnh $\upsilon(\boldsymbol{X})$ trù mật trong [0,1].
- $\upsilon(x)$ chính là giá trị đầu mút chung của hai khoảng $\mathfrak{I}(h_{-1}x)$ và $\mathfrak{I}(h_1x)$ và nó là điểm phân tách sao cho các khoảng $\mathfrak{I}(hx)$ với gia tử h mang cùng dấu chỉ ở một phía của $\upsilon(x)$. Giá trị $\upsilon(x)$ được xác định như vậy hoàn toàn tính được bằng một công thức đệ quy (xem chẳng hạn [6]).

(iv) Khoảng tương tự mức k: Đặt $\boldsymbol{X}_{(k)} = \{x \in \boldsymbol{X}: |x| \le k\}$ gồm các từ có độ dài không lớn hơn k. Bài toán đặt ra là phân hoạch đoạn [0,1] thành các khoảng rời nhau sao cho mỗi khoảng của phân hoạch chứa duy nhất một giá trị định lượng $\upsilon(x)$ của $\boldsymbol{X}_{(k)}$. Phân hoạch như vậy sẽ xác định một quan hệ tương tự S (similarity) theo nghĩa Blookles-Pety [5] hay Shenoi [15]. Vì vậy, nó phản ánh một ngữ nghĩa là mỗi khoảng của phân hoạch mà chứa $\upsilon(x)$ nào đó sẽ bao gồm các giá trị tương tự với $\upsilon(x)$, theo quan hệ S.

Bài toán này được giải theo nghĩa topo như sau: Vì họ các khoảng tính mờ của $\boldsymbol{X}$ lập thành một cơ sở topo của $\boldsymbol{X}$, nghĩa là nó xác định một topo trên [0,1] sao cho mỗi tập mở của [0,1] đều là hợp của một số khoảng tính mờ. Xét các khoảng tính mờ của $\boldsymbol{X}_{k+2}$ và phân hoạch các khoảng này thành các lớp $C_k(x)$, $x \in \boldsymbol{X}_{(k)}$, sao cho chúng chứa ít nhất 2 khoảng tính mờ của $\boldsymbol{X}_{k+2}$ mà đầu mút chung của chúng là giá trị định lượng $\upsilon(x)$. Đặt $S_k(x) = \bigcup\{\mathfrak{I}_{k+1}: \mathfrak{I}_{k+1} \in C_k(x)\}$. Họ $\{S_k(x): x \in \boldsymbol{X}_{(k)}\}$ là một phân hoạch của [0,1] và mỗi $S_k(x)$ được gọi là *khoảng tương tự mức k* của $x \in \boldsymbol{X}_{(k)}$ của quan hệ tương tự mức k, kí hiệu là S_k (xem [4]). Tóm lại, các khoảng tương tự như vậy có tính chất sau:

(i) $\{S_k(x): x \in \boldsymbol{X}_{(k)}\}$ lập thành phân hoạch của đoạn [0,1].

(ii) $\upsilon(x) \in S_k(x)$ và là đầu mút chung của ít nhất 2 khoảng của $\boldsymbol{X}_{k+2}$.

Định nghĩa 3.1 [11]. *Cho các giá trị tham số tính mờ, biểu diễn bộ 4 của* $x \in \boldsymbol{X}_{(l)}$ là

$$\textit{bộ 4}: (x, \upsilon(x), r, S_l(x)), \text{ với } r \in S_l(x)$$

trong đó $\upsilon(x)$ là giá trị định lượng của từ x, $S_l(x)$ là khoảng tương tự ngữ nghĩa mức l của x, gọi là mức tương tự. Ta luôn luôn có $\upsilon(x) \in S_l(x)$ và ý nghĩa của $\upsilon(x)$ giống như là tâm (*core*) của một tập mờ, nghĩa là nó là giá trị tương hợp nhất với ngữ nghĩa của từ x. ■

Giá trị r xuất hiện ở thành phần thứ 3 trong biểu diễn bộ 4 với ba mục tiêu:

- Ta sẽ dùng các giá trị ngữ nghĩa định lượng để kết nhập ý kiến và kết quả sẽ là một giá trị thực r nói chung không trùng với bất kì giá trị định lượng nào của các từ trong thang điểm. Vì vậy cần dành thành phần của bộ 4 cho nó, thành phần thứ 3, để lưu.

- Mở rộng để linh hoạt, ta giả định rằng một ứng dụng cho phép các chuyên gia có thể vừa biểu thị đánh giá bằng từ ngôn ngữ x, vừa biểu thị ý kiến bằng điểm số r. Trường hợp chuyên gia cho đánh giá bằng từ ngôn ngữ x, ý kiến đó sẽ được biểu diễn bằng bộ 4 $(x, \upsilon(x), \upsilon(x), S_l(x))$. Trong những trường hợp đánh giá bằng giá trị $r \in [0,1]$, nó cũng được biểu diễn bằng bộ 4 $(x, \upsilon(x), r, S_l(x))$, trong đó $S_l(x)$ là khoảng tương tự duy nhất chứa r. Vì thành phần r xác định duy nhất các thành phần còn lại của bộ 4 nên để tiện dụng nó được kí hiệu là $(x(r), \upsilon(x(r)), r, S_l(x(r)))$.

- Trường hợp trong một data warehouse có chứa các điểm đánh giá các đối tượng (sinh viên, học sinh, các chuyên gia công nghệ, …) có thể xảy ra điểm đánh giá vừa biểu thị bằng số, vừa bằng từ ngôn ngữ.

Hiệu số $r - \upsilon(x)$ được gọi là độ lệch phù hợp ngữ nghĩa của giá trị r.

Ta phát biểu mệnh đề sau đây để tiện dụng sau này.

Mệnh đề 3.1. Nếu $r \le r'$ thì các từ ngôn ngữ $x(r)$ và $x(r')$ trong biểu diễn bộ 4 của r và r' thỏa bất đẳng thức $x(r) \le x(r')$.

Chứng minh: Chứng minh của mệnh đề này rút ra trực tiếp từ Định nghĩa 3.1. Thực vậy, vì các khoảng $I_l(x)$, $x \in \boldsymbol{S}$, lập thành một phân hoạch của đoạn [0,1], ta thấy hoặc có một khoảng $I_l(x)$ sao cho $r, r' \in I_l(x)$, hoặc là $r \in I_l(x)$ và $r' \in I_l(x')$ với $I_l(x) \ne I_l(x')$. Trong trường hợp thứ nhất ta có $x(r) = x(r')$ và trong trường hợp thứ 2 ta suy ra $I_l(x) < I_l(x')$ và do đó $x(r) < x(r')$. ■

Định nghĩa 3.2. *Thang điểm ngôn ngữ bộ 4* bao gồm các giá trị ngôn ngữ bộ 4 sau:

$$\{(x, \upsilon(x), r, S_l(x)): x \in \boldsymbol{S}, r \in [0,1]\}.$$ ■

Biểu diễn này có thể được xem như thỏa mãn Nguyên tắc 2 vì nó mang nhiều thông tin ngữ nghĩa định tính dựa trên các đặc trưng lượng hóa phong phú của ĐSGT. Những thông tin này phong phú hơn nhiều so với cách tiếp cận dựa trên biểu diễn bộ 2 và cách tiếp cận dựa trên toán tử LOWA được trình bày trong Mục II.

➢ *Xây dựng thang điểm ngôn ngữ bộ 4*:

Tập $\boldsymbol{S}$ các từ ngôn ngữ trong cách tiếp cận này được xác định là tập con của ĐSGT thỏa mãn điều kiện nào đấy. Trong [4] giả thiết thang điểm $\boldsymbol{S}$ là *đầy*, tức là tồn tại k sao cho $\boldsymbol{S} = \boldsymbol{X}_{(k)}$. Việc này hạn chế linh hoạt trong ứng dụng, chẳng hạn thang điểm $\boldsymbol{S}$ = {*bad*, *medium*, *Rather_good*, *good*, *Very_good*} $\ne \boldsymbol{X}_{(2)}$, vì nó không chứa *Very_bad* và *Rather_bad*.

Ta giả thiết $\boldsymbol{S}$ thỏa điều kiện: $(x = hx' \in \boldsymbol{S}) \Rightarrow (x' \in \boldsymbol{S})$. Ví dụ, nếu *Very_bad* $\in \boldsymbol{S}$ thì *bad* $\in \boldsymbol{S}$. Điều kiện này là tự nhiên trong thực tế, vì nếu có phân loại học sinh *rất kém* thì phải có tiêu chí phân loại *kém*.

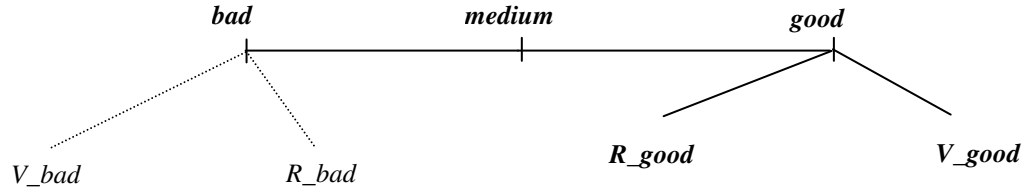
Để xây dựng biểu diễn thang điểm ngôn ngữ $\boldsymbol{S}$, ta tiến hành các bước sau:

Bước 1: Xác định độ dài từ lớn nhất l của $\boldsymbol{S}$ và xác định một cách trực giác các giá trị tham số tính mờ của $\boldsymbol{S}$. Tính các khoảng tương tự mức l, $S_l(s)$, của các từ trong $\boldsymbol{X}_{(l)}$.

Đặt $I_l(x) := S_l(x), x \in \boldsymbol{X}_{(l)}$.

Bước 2: Thực hiện quy nạp theo $k = l - |x|$, $x \in \boldsymbol{S}$, thao tác tính $I_l(x)$: Giả sử đối với tất cả các từ $x \notin \boldsymbol{S}$ với $l - |x| < k$, khoảng $I_l(x)$ đã được tính. Khi đó ta lần lượt xét $x = hx' \in \boldsymbol{X}_{(l)}$ với $l - |x'| = k$, nếu $x \notin \boldsymbol{S}$ thì tính $I_l(x') := I_l(x') \cup S_l(x')$, cho đến khi không còn x như vậy.

Ví dụ 3.1. Để minh họa, xét ví dụ tập $\boldsymbol{S}$ ở trên. Nó có biểu diễn dạng cây như Hình 1.



Hình 1. Cây biểu diễn tập $\boldsymbol{S}$

Ta thấy $l = 2$ và chỉ có 2 từ *V_bad* và *R_bad* là không thuộc $\boldsymbol{S}$. Lần lượt xét $x_1 = V\_bad$ và $x_2 = R\_bad$ và $I_2(bad)$ lần lượt được tính là

$I_2(bad) := S_2(bad) \cup S_2(V\_bad)$ và do đó $I_2(bad) := S_2(bad) \cup S_2(V\_bad) \cup S_2(R\_bad)$.

Với các x còn lại, $I_2(x) = S_2(x)$. Kết quả ta thu được thang điểm biểu diễn bộ 4 có dạng:

$$\{(x, \upsilon(x), r, I_2(x)) : x \in \boldsymbol{S}, r \in [0,1]\} \quad \blacksquare$$

➢ *So sánh giữa các giá trị trong thang điểm bộ 4*: Yếu tố cốt yếu trong việc lấy quyết định là các phương án lựa chọn cần được sắp xếp thứ tự ưu tiên. Vì vậy, các giá trị trong thang điểm đòi hỏi phải sánh được với nhau.

Định nghĩa 3.3. Cho hai điểm ngôn ngữ biểu diễn bộ 4 $(x, \upsilon(x), r, I_l(x))$ và $(x', \upsilon(x'), r', I_l(x'))$. Ta định nghĩa: $(x, \upsilon(x), r, I_l(x)) \le (x', \upsilon(x'), r', I_l(x'))$ khi và chỉ khi một trong hai điều kiện sau đúng:

(i) $x \le x'$.

(ii) $x = x'$ và $r \le r'$. $\blacksquare$

Có thể dễ dàng thấy thang điểm ngôn ngữ bộ 4 là thang điểm sắp toàn phần (tuyến tính). Về ý nghĩa của sự sắp xếp này là tự nhiên, phù hợp với trực quan của chúng ta.

3) Việc kết nhập các ý kiến đánh giá trên thang ngôn ngữ bộ 4 và Nguyên tắc 3

Xét thang điểm ngôn ngữ biểu diễn ngữ nghĩa bộ 4 $\{(x, \upsilon(x), r, I_l(x)) : x \in \boldsymbol{S}, r \in [0,1]\}$. Trong tiếp cận này, thành phần thứ ba r sẽ được sử dụng trong tính toán kết nhập các ý kiến chuyên gia hay kết nhập các điểm đánh giá theo các tiêu chuẩn. Những giá trị r như vậy trong biểu diễn bộ 4 mang nhiều thông tin:

(i) $r \in I_l(x)$ có nghĩa r phù hợp với ngữ nghĩa của từ x. Chẳng hạn x là từ "*xuất sắc*", Khoảng $I_l(x)$ là khoảng điểm được xếp là *xuất sắc*.

(ii) Nếu một chuyên gia cho điểm bằng giá trị ngôn ngữ x thì thành phần thứ ba r sẽ là $\upsilon(x)$ - giá trị ngữ nghĩa định lượng của x. Khi đó r là giá trị *tâm* (*core*) của x hay nó mang nhiều ngữ nghĩa nhất về x.

Vì vậy, việc kết nhập các thành phần thứ ba của các điểm đánh giá bộ 4 tương tự như việc đánh giá và phân loại đối tượng (thành tích học tập của sinh viên chẳng hạn) theo các nhóm với nhãn đánh giá ngôn ngữ (chẳng hạn, sinh viên được xếp loại *xuất sắc*).

Định nghĩa 3.4. Giả sử α là phép kết nhập các giá trị của đoạn [0,1] theo nghĩa Yager. Khi đó, α sẽ cảm sinh một phép kết nhập Agg_α trên các điểm bộ 4 được xác định như sau: với mỗi vec tơ các giá trị bộ 4 $\boldsymbol{a} = (a_1, \ldots, a_n)$, $a_i = (x_i, \upsilon(x_i), r_i, I_l(x_i))$, $i = 1, \ldots, n$, ta định nghĩa

$$Agg_\alpha(\boldsymbol{a}) = Agg_\alpha(a_1, \ldots, a_n) = (x(r_a), \upsilon(x(r_a), r_a, I_l(x(r_a))$$

trong đó $r_a = \alpha(r_1, \ldots, r_n)$. ■

Bây giờ chúng ta sẽ chứng tỏ rằng Agg_α thỏa mãn đòi hỏi về tính đơn điệu tăng thực sự trong *Nguyên tắc 3*:

Mệnh đề 3.2. Agg_α là phép toán đơn điệu tăng và nếu a là phép kết nhập số đơn điệu tăng thực sự thì Agg_α cũng vậy.

Chứng minh: Xét hai vec tơ các giá trị bộ 4 của thang điểm ngôn ngữ bộ 4 gồm l từ ngôn ngữ $\boldsymbol{a}$ và $\boldsymbol{b}$ với các thành phần $a_i = (x_i, \upsilon(x_i), r_{a,i}, I_l(x_i))$ và $b_i = (y_i, \upsilon(y_i), r_{b,i}, I_l(y_i))$, $i = 1, \ldots, n$, và giả sử $\boldsymbol{a} \le \boldsymbol{b}$.

Ta chứng tỏ rằng $\boldsymbol{a} \le \boldsymbol{b}$ kéo theo $r_{a,i} \le r_{b,i}$, $i = 1, \ldots, n$. Thực vậy, với $a_i \le b_i$ có hai khả năng xảy ra. Một là $x_i < y_i$. Vì các khoảng $I_l(x_i)$, $i = 1, \ldots, n$, lập thành phân hoạch của đoạn [0,1] và $\upsilon(x_i) \le \upsilon(y_i)$, ta suy ra $I_l(x_i) < I_l(y_i)$. Theo ràng buộc đối với thành phần thứ ba của bộ 4 ta có $r_{a,i} < r_{b,i}$. Hai là $x_i = y_i$ và do vậy theo định nghĩa về quan hệ so sánh $a_i \le b_i$ ta phải có $r_{a,i} \le r_{b,i}$. Vì α đơn điệu tăng, ta có $r_a = \alpha(r_{a,1}, \ldots, r_{a,n}) \le r_b = \alpha(r_{b,1}, \ldots, r_{b,n})$. Theo Mệnh đề 3.1, Định nghĩa 3.3 và 3.4, suy ra $Agg_\alpha(\boldsymbol{a}) \le Agg_\alpha(\boldsymbol{b})$.

Bây giờ giả sử $\boldsymbol{a} < \boldsymbol{b}$. Khi đó có ít nhất một thành phần i sao cho $(x_i, \upsilon(x_i), r_{a,i}, I_l(x_i)) < (y_i, \upsilon(y_i), r_{b,i}, I_l(y_i))$. Chứng minh tương tự như trên ta có hoặc $x_i < y_i$, hoặc $r_{a,i} < r_{b,i}$. Vì phép kết nhập α đơn điệu tăng thực sự, ta phải có $r_a = \alpha(r_{a,1}, \ldots, r_{a,n}) < r_b = \alpha(r_{b,1}, \ldots, r_{b,n})$. Từ đó, theo Định nghĩa 3.4 ta suy ra $Agg_\alpha(\boldsymbol{a}) < Agg_\alpha(\boldsymbol{b})$, nghĩa là Agg_α đơn điệu tăng thực sự. ■

Như vậy *thang điểm ngôn ngữ bộ 4* thỏa mãn 3 nguyên tắc đề ra trong Mục 2.

4. VAI TRÒ CỦA THANG ĐIỂM BIỂU DIỄN BỘ 4 TRONG VẤN ĐỀ LẤY QUYẾT ĐỊNH

Trong [4], đã khảo sát kết quả của việt lấy quyết định dựa trên biểu diễn bộ 4, nhưng ví dụ chỉ chỉ ra giá trị định lượng khác nhau so với cách tiếp cận bộ 2, nhưng phương án được lựa chọn cho người ra quyết định vẫn giống nhau, mặc dù biểu diễn bộ 4 có cơ sở ngữ nghĩa định lượng có bản chất phù hợp hơn, và vì vậy đúng đắn hơn. Điều chúng ta kì vọng là cần phải chứng tỏ các kết quả lấy quyết định là khác nhau đối với hai cách tiếp cận dựa trên bộ 2 và bộ 4, và tất nhiên kết quả quyết định có cơ sở phương pháp luận tốt hơn sẽ có độ tin cậy cao hơn.

Nhằm mục đích này, ta chỉ cần chỉ ra ví dụ rằng hai cách tiếp cận được đề cập sẽ cho kết quả đánh giá làm thay đổi thứ tự sắp xếp 2 phương án lựa chọn quyết định.

Xét thang điểm có 9 từ ngôn ngữ sắp xếp thừ nhỏ đến lớn:

$\boldsymbol{S} = \{s_i : i = 1, \ldots, 9\} = \{E\_bad := \boldsymbol{0}, V\_bad, bad, R\_bad, medium, R\_good, good, V\_good, E\_good := \boldsymbol{1}\}$

trong đó ***0*** và ***1*** là hai khái niệm ngôn ngữ nhỏ nhất và lớn nhất được mã hóa là *E_bad* và *E_good* (*E* := *Extreamly*) của ĐSGT gồm 2 gia tử *R* (*rather*) và *V* (*very*) được gán cho thang điểm ngôn ngữ của 3 tiêu chuẩn: C_1 := giải pháp công nghệ. C_2 := phương án thực hiện và C_3 := giải pháp tài chính.

Thang điểm $\boldsymbol{S}$ đã cho là tập $\boldsymbol{X}_{(2)} \cup \boldsymbol{C}$, tức là gồm các từ độ dài không lớn hơn 2 và các hằng ***0***, ***W*** và ***1***. Vậy, để xây dựng các khoảng tương tự $S_2(s_i)$ của thang điểm chúng ta cần sử dụng *khoảng tính mờ* của tập $\boldsymbol{X}_4$ gồm các từ độ dài 4 được liệt kê tăng dần từ trái sang phải như sau:

$$VVVc^-, RVVc^-, RRVc^-, VRVc^-, VRRc^-, RRRc^-, RVRc^-, VVRc^-,$$

$$VVR\ c^+, RVRc^+, RRRc^+, VRRc^+, VRVc^+, RRVc^+, RVVc^+, VVVc^+.$$

Lưu ý rằng các xâu gia tử trong dãy 16 từ này đối xứng gương khi chuyển từ c^- sang c^+. Trừ các khoảng tính mờ của từ đầu và cuối trong 16 từ ở trên, tương ứng sẽ là khoảng tương tự của *E_bad* và *E_good*, hợp của 2 khoảng tính mờ của hai từ kế tiếp tương ứng sẽ là khoảng tương tự mức 2 của 7 từ còn lại trong thang điểm.

Để xác định ngữ nghĩa hợp với thực tế ta chọn các giá trị tham số sao cho khoảng điểm số ứng với từ *médium* có cận trái là 0,5. Từ đó ta có hệ thức:

$$fm(c^-) = 0{,}5/\mu^2(V)(1 - \mu(V))$$

Như ta đã biết, các đặc trưng lượng hóa của ĐSGT chỉ phụ thuộc vào các tham số tính mờ. vì vậy khéo lựa chọn tham số tính mờ $\mu(V)$, cụ thể $\mu(V) = 0{,}484$. Do đó, theo hệ thức trên ta có $fm(c^-) = 0{,}5687$ và ta tính được được các kết quả trong bảng 1a và 1b.

Bảng 1[a]. Độ đo tính mờ của các từ trong X(4) với từ nguyên thủy âm c–

VVVc-	*LVVc-*	*LLVc-*	*VLVc-*	*VLLc-*	*LLLc-*	*LVLc-*	*VVLc-*
0,0645	*0,0687*	*0,0733*	*0,0687*	*0,0733*	*0,0781*	*0,0733*	*0,0687*

Bảng 1b. Độ đo tính mờ của các từ trong X(4) với từ nguyên thủy âm c+

VVLc+	*LVLc+*	*LLLc+*	*VLLc+*	*VLVc+*	*LLVc+*	*LVVc+*	*VVVc+*
0,052128	*0,055575*	*0,059249*	*0,055575*	*0,052128*	*0,055575*	*0,052128*	*0,048895*

Từ các dữ kiện này, biểu diễn bộ 4 của thang điểm ngôn ngữ trên miền thang điểm 10 được xác định như sau:

(*E_bad*, 0,31, r_1, [0, 0,65)) (*V_bad*, 1,33, r_2, [0,65, 2,07))

(*bad*, 2,75, r_3, [2,07, 3,49)) (*R_bad*, 4,27, r_4, [3,49, 0,5))

(*médium*, 5,69, r_5, [0,5, 6,21)) (*R_good*, 6,77, r_6, [6,21, 7,36))

(*good*, 7,91, r_7, [7,36, 8,43)) (*V_good*, 8,99, r_8, [8,43, 9,51))

(*E_good*, 9,76, r_9, [9,51, 1))

➢ *So sánh các phương pháp tiếp cận dựa trên bài toán lấy quyết định đa tiêu chuẩn*

Bài toán lấy quyết định đa tiêu chuẩn theo nhóm chuyên gia có thể tách ra để quy về mô hình hình thức hóa lấy quyết định chỉ theo đa tiêu chuẩn hoặc lấy quyết định chỉ theo nhóm chuyên gia. Vì vậy, trong nghiên cứu này chúng ta có thể nghiên cứu ví dụ giải bài toán lấy quyết định đa tiêu chuẩn để phân tích kết quả theo 2 cách tiếp cận, biểu diễn bộ 2 và biểu diễn bộ 4, để so sánh. Về mặt phương pháp luận sẽ không có gì khó khăn nếu đòi hỏi thêm việc lấy quyết định theo nhóm chuyên gia, vì khi đó ta chỉ cần thực hiện thêm việc kết nhập.

Chúng ta xét hai phương án dự thầu A_1 và A_2 được một chuyên gia đánh giá điểm ngôn ngữ theo 3 tiêu chuẩn C_1, C_2 và C_3 với trọng số khác nhau như biểu thị trong bảng 2.

Bảng 3. Điểm ngôn ngữ của ban giám khảo theo từng tiêu chuẩn với trọng số đã cho

Tiêu chuẩn C và trọng số w	*C1, w1 = 0,25*	*C2, w2 = 0,5*	*C3, w3 = 0,25*
A1	*s9*	*s2*	*s9*
A2	*s6*	*s4*	*s7*

➢ *Theo cách tiếp cận biểu diễn từ ngôn ngữ theo bộ 2*

Theo cách tiếp cận này việc kết nhập các điểm ngôn ngữ được tính trên các chỉ số của các điểm ngôn ngữ. Kết quả kết nhập theo 3 tiêu chuẩn được cho trong nửa trên của bảng 4. Như vậy, điểm của phương án A_1 là bộ 2 (l_6, -0,5), hay (*R_good*, -0,5), còn điểm của phương án A_2 là (l_5, +0,25), hay (*medium*, +0,25). Vậy, A_1 được khuyến cáo cho người lấy quyết định nên chọn.

➢ *Theo cách tiếp cận biểu diễn từ ngôn ngữ theo bộ 4*

Như đã trình bày ở trên, trong cách tiếp cận này chuyên gia có thể cho điểm theo từ ngôn ngữ hoặc theo điểm số. Về mặt kĩ thuật, việc này rất thuận tiện vì khi tính toán kết quả cho điểm sau kết nhập luôn luôn là con số. Trong thực tiễn, việc này cũng cho phép phương pháp luận trở nên linh hoạt, chẳng hạn khi ta phân tích các kết quả đánh giá các đối tượng trong một lĩnh vực quản lí trong một thời gian dài ta sẽ gặp các dữ kiện đánh giá theo thang điểm số trước kia và theo điểm ngữ sau này. Cách tiếp cận dựa trên bộ 4 cho phép biểu diễn thống nhất hai cách cho điểm như vậy. Kết quả tính toán kết nhập theo thành phần thứ 3 của biểu diễn bộ 4 của các điểm đánh giá của chuyên gia được ghi nhận trong nửa dưới của bảng 4.

Như vậy, điểm kết nhập trong thang điểm ngôn ngữ bộ 4 của phương án A_1 là (*medium*, 5,69, 5,55, [0,5, 6,21)), còn của phương án A_2 là (*medium*, 5,59, 5,80, [0,5, 6,21)). Khác với cách tiếp cận trên, phương án được khuyến cáo là A_2 vì 5,80 > 5,59.

Với cách tiếp cận bộ 4, người lấy quyết định có thể rút ra nhiều thông tin từ kết quả tính toán so với cách tiếp cận bộ 2:

- Một thông tin quan trọng là ngữ nghĩa định lượng của từ ngôn ngữ. Nó mang ý nghĩa như lõi (core) của tập mờ: nó là giá trị số tương hợp nhất với ngữ nghĩa của từ ngôn ngữ đang xét, nó là giá trị mốc để *áng chừng* đánh giá xem các đánh giá điểm bằng giá trị số hay giá trị ngôn ngữ khác sẽ khác bao xa so với điểm ngôn ngữ đang xét. Trong khi đó giá trị chỉ số trong cách tiếp cận bộ 2 mang rất ít thông tin: nó chỉ mang thông tin thứ tự giữa các từ mà không phản ảnh đúng ngữ nghĩa định lượng của chúng. Chẳng hạn, khi số lượng các từ ngôn ngữ trong ví dụ trên thay đổi từ 9 xuống còn 5 bằng việc loại bỏ 4 từ có thứ tự nhỏ, khi đó các kết quả tính toán

đối với cách tiếp cận bộ 2 sẽ thay đổi lớn, nhưng đối với cách tiếp cận bộ 4 thì không. Điều này chỉ ra tính không hợp lí của cách tiếp cận bộ 2.

- Dữ liệu về khoảng tính mờ trong biểu diễn bộ 4 chỉ khoảng điểm số ứng với nó, cũng cung cấp cho người quyết định nhiều thông tin tham khảo. Nếu khoảng này rộng, nó nói rằng sự khác biệt giữa các điểm đánh giá là trong khoảng là đáng kể, hay các đối tượng được xếp vào cùng một nhóm theo tiêu chuẩn này có thể có một khoảng cách điểm đáng kể được đo bằng đại lượng $(r - \upsilon(s))$. Chẳng hạn, khoảng điểm số của *E_good* là [9,51, 1) có độ dài là 0,49, khá rất nhỏ so với hầu hết độ dài các khoảng khác, chỉ ra rằng điểm số của những đối tượng cùng suất sắc theo tiêu chuẩn này là không lớn. Nó phù hợp với thực tế mong muốn là số các đối tượng *E_good* là ít.

Bảng 4. Kết quả tính kết nhập điểm theo 3 tiêu chuẩn có trọng số

	C_1	C_2	C_3	*KQ kết nhập*
	$w_1 = 0,25$	$w_2 = 0,5$	$w_3 = 0,25$	
Theo cách tiếp cận biểu diễn bộ 2				
A_1	9	2	9	5.5
A_2	6	4	7	5.25
Theo cách tiếp cận biểu diễn bộ 4				
A_1	9.76	1.33	9.76	5.55
A_2	6.76	4.27	7.91	5.80

5. KẾT LUẬN

Trong nghiên cứu này chúng tôi chỉ ra rằng quan hệ lôgic giữa cú pháp (kí hiệu) và ngữ nghĩa là rất quan trọng trong nghiên cứu lôgic và thông tin mờ nói chung và trong nghiên cứu bài toán lấy quyết định mờ nói riêng. Quan hệ này có tính phổ quát trong lĩnh vực nghiên cứu công nghệ thông tin. Nghĩa là, ở mức độ khác nhau vấn đề này cần quan tâm thích đáng.

Việc phân tích những nhược điểm của các cách tiếp cận khác, kể cả cách tiếp cận bằng tập mờ, trong đó tập mờ không phản ánh ngữ nghĩa thông qua quan hệ thứ tự vốn có của ngôn ngữ, chứng tỏ rằng việc không quan tâm đến vấn đề cú pháp và ngữ nghĩa đang hiện diện trong nhiều nghiên cứu.

Thiếu sót này chỉ ra ưu thế của ĐSGT vì nó phản ánh phù hợp nhất mối quan hệ cú pháp và ngữ nghĩa của các từ ngôn ngữ mờ. Đặc biệt nó phản ánh phù hợp với thức tiễn khi hình thành thang điểm. Ta hãy quan sát lại thang điểm ngôn ngữ trong ví dụ nghiên cứu trong Mục 4, nếu hợp 3 khoảng điểm của ba từ ngôn ngữ đầu tiên của thang điểm lại với nhau, ta thu được khoảng [0, 3,49) mà nên gán nó với điểm ngôn ngữ *bad* (*kém*), tương đối phù hợp với khoảng điểm thực tế đang sử dụng ở trường phổ thông khi đánh giá kết quả từng môn học. Khi đó các từ ngôn ngữ của thang điểm tương đồng với các từ phân loại học sinh theo môn học: *kém*, *yếu*, *trung bình*, *khá*, *giỏi*, *suất sắc*, và các khoảng điểm tương ứng của chúng cũng khá phù hợp.

Việc xây dựng ví dụ về phương pháp lấy quyết định trong Mục 4 chỉ ra kết quả khuyến cáo lấy quyết định khác nhau đối với hai phương án lựa chọn không đơn thuần chỉ là kĩ thuật xây

dựng, mà nó phản ánh bản chất giải quyết quan hệ cú pháp và ngữ nghĩa khác nhau, là hệ quả tất yếu của sự khác biệt trong biểu diễn ngữ nghĩa của các từ ngôn ngữ mờ. Chúng tôi cho rằng biểu diễn bộ 4 của thang điểm ngôn ngữ có cơ sở phương pháp luận đúng đắn hơn và, do đó, nó cho kết quả tin cậy hơn.

Lời cảm ơn. Bài báo được sự tài trợ của Quỹ Phát triển khoa học Công nghệ Quốc gia Nafosted, mã số 102.01- 2011.06

TÀI LIỆU THAM KHẢO

1. Buckles B.P., Petry F.E. - A Fuzzy representation of data for Relational Databases, Fuzzy Sets and Systems **7** (3) (1982) 213-226
2. Herrera F. and Martinez L. - A Model Based on Linguistic 2-Tuples for Dealing with Multigranular Hierachical Linguistic Contexts in Multi-Expert Decision-Making, IEEE trans. on syst., man, and cyber. **31** (2) (2001), 227-234.
3. Herrera F., Herrera-Viedma E. - Linguistic decision analysis: steps for solving decision problems under linguistic information, Fuzzy Sets and Syst **115** (2000) 67-82.
4. Long V. Ng., Thông V. Ng. - Vấn đề kết nhập thông tin biểu diễn bằng bộ 4 với ngữ nghĩa dựa trên đại số gia tử, Tạp chí Tin học và Điều khiển học **27** (3) (2011) 241-252.
5. Delgado M., J. Verdegay L., and Vila M. A. - On aggregation operations of linguistic labels, Int. J. Intell. Syst. **8** (1993) 351-370.
6. Herrera F. and Martinez L. - A Model Based on Linguistic 2-Tuples for Dealing with Multigranular Hierachical Linguistic Contexts in Multi-Expert Decision-Making, IEEE trans. on syst., man, and cyber. **31** (2) (2001) 227-234.
7. Ho C Ng., Long N. V. - Fuzziness Measure on Complete Hedge Algebras and Quantitative Semantics of Terms in Linear Hedge Algebras, Fuz. Sets and Syst. **158** (2007) 452-471.
8. Hung V. Le, Ho C. Ng., Fei Liu. - Semantics and Aggregation of Linguistic Information Based on Hedge Algebras, The 3th Inter. Conf. on Knowledge, Infor. and Creativity Support Syst., KICSS 2008, Hanoi, Vietnam, Dec. 22-23, 2008, pp. 128-135.
9. Nam V. Huynh, Ho C. Ng., Nakamori Y. - MEDM in General Multi-granular Hierarchical Linguistic Contexts Based on The 2-Tuples Linguistic Model, In: Proc. of IEEE Int. Conf. on Granular Computing, 2005, pp. 482-487.
10. Jun Liu, Da Ruan, Roland Carchon. - Synthesis and Evaluation Analysis of the Indicator Information in Nuclear Safeguards Applications by Computing Withwords, Int. J. Appl. Math. Comput. Sci. **12** (3) (2002) 449-462.
11. Long V. Ng. - Cơ sở phương pháp luận của phương pháp đánh giá bằng nhãn ngôn ngữ, Tạp chí Tin học và Điều khiển học **24** (1) (2008) 75-86.
12. Ho C. Ng., Nam V. Huynh - Ordered Structure-Based Semantics of Linguistic Terms of Linguistic Variables and Approximate Reasoning, AIP Conf. Proceed. on Computing Anticipatory Systems, CASYS'99, 3th Inter. Conf., 1999, pp. 98-116.
13. Ho Cat Ng., Nam V. Huynh. - Towards an Algebraic Foundation for a Zadeh Fuzzy Logic, Fuzzy Set and System **129** (2002) 229-254.

14. Cát Hồ Ng., Sơn Th. Tr., Khang Đ. Tr., Việt X. Lê - Fuzziness Measure, Quantified Semantic Mapping And Interpolative Method of Approximate Reasoning in Medical Expert Systems, Journal of Informatics and Cybernetics **18** (3) (2002) 237-252.
15. Shenoi S., Melton A. - Proximity relations in the fuzzy relational database model, Fuzzy Sets and Systems **31** (1989) 285-296.
16. Yager R. - On ordered weighted averaging aggregation operators in multicriteria decisionmaking, IEEE Trans. on Systems, Man and Cybernetics **18** (1) (1988) 183–190.
17. Yager R., Rybakov A. - Uniform Aggregation Operators, Fuzzy sets and Systems **80** (1996) 111-120.
18. Yager R., Rybakov A. - Non-commutative Self-identity Aggregation, Fuzzy sets and Systems **85** (1997) 73-82.

ABSTRACT

ON RELATION BETWEEN SYNTAX AND SEMANTICS OF TERMS IN BUILDING METHODS FOR SOLVING MULTI-CRITERIA DECISION MAKING PROBLEMS

Nguyễn Cát Hồ[1], Nguyễn Thanh Thủy[2], Nguyễn Văn Long[3], Lê Ngọc Hưng[4,*]

[1]*Institute of Information Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Hanoi, Vietnam*

[2]*University of Engineering and Technology (VNU), 44 Xuan Thuy, Cau Giay, Hanoi, Vietnam*

[3]*University of Transport and Communications, Hanoi, Vietnam*

[4]*Sai Gon University, 273 An Duong Vuong, Ho Chi Minh city, Vietnam*

[*]Email: *lengochunsg291958@gmail.com*

In some recent approach the meaning of term is interpreted as their indexes in a linguistic scale, which are not stable and depend on the position and the cardinality of the scale. A semantic approach to construct linguistic scales based on the examination of the logical relation between syntax and semantics of terms is proposed. Some principles are introduced for such a construction to solve multi-criteria decision making (MCDM) problems, including a criterion for aggregation operator. It is shown that the approach based on 4-tuple representation can be considered to follow these discussed principles well. Examples of solving some MCDM problems for illustration are examined to show the soundness of the proposed approach.

Keywords: linguistic scale, multi-criteria decision making problem, hedge algebra, interval semantics, aggregation operator.