**BÀI BÁO**

*Một hàm niềm tin cơ sở mở rộng trong lý thuyết bằng chứng Dempster-Shafer và nó ứng dụng trong hợp nhất dữ liệu xung đột*

**Tóm lược:** Lý thuyết bằng chứng Dempster–Shafer đã được áp dụng rộng rãi trong lĩnh vực hợp nhất thông tin. Tuy nhiên, khi dữ liệu bằng chứng được thu thập có tính xung đột cao, quy tắc kết hợp Dempster (Dempster combination rule - DCR) không tạo ra kết quả trực quan trong hầu hết thời gian. Để giải quyết vấn đề này, cơ sở hàm niềm tin được đề xuất để sửa đổi phép gán xác suất cơ bản (basic probability assignment - BPA) trong khung đầy đủ nhận thức (frame of discernment - FOD). Tuy nhiên, trong FOD không đầy đủ, giá trị hàm khối lượng của tập rỗng khác không, điều này làm cho hàm niềm tin cơ sở không còn áp dụng được nữa. Trong bài báo này, xem xét các ảnh hưởng của kích thước của FOD và giá trị hàm khối lượng của tập rỗng, một hàm niềm tin mới được đặt tên là hàm niềm tin cơ sở mở rộng (extended base belief function - EBBF) được đề xuất. Phương pháp này có thể sửa đổi BPA trong FOD không đầy đủ và thu được kết quả hợp nhất trực quan bằng cách tính đến các đặc điểm của FOD không đầy đủ. Ngoài ra, EBBF có thể suy biến thành hàm niềm tin cơ sở trong FOD đầy đủ. Đồng thời, bằng cách tính toán entropy niềm tin của BPA đã sửa đổi, chúng tôi thấy rằng giá trị của niềm tin entropy cao hơn trước. Niềm tin entropy được sử dụng để đo lường sự không chắc chắn thông tin, có thể hiển thị xung đột một cách trực quan hơn. Sự gia tăng giá trị của entropy niềm tin là hệ quả của xung đột. This paper also designs an improved conflict data management method based on the EBBF to verify the rationality and effectiveness of the proposed method.

**Từ khóa:**

* Lý thuyết bằng chứng Dempster–Shafer (D-S)
* Chức năng niềm tin
* Khung không đầy đủ của nhận thức (FOD)
* Chức năng niềm tin cơ sở
* Bằng chứng mâu thuẫn
* Quy tắc kết hợp tổng quát

1. **Giới thiệu**

Với sự phát triển của công nghệ máy tính, Internet và các lĩnh vực liên quan khác, công nghệ tổng hợp thông tin, được sinh ra trong lĩnh vực quân sự [1], đã chạy qua mọi ngóc ngách của sản xuất và đời sống của nhân dân [2]. Việc thực hiện hợp nhất thông tin cần phải đối phó với rất nhiều thông tin không chắc chắn. Các công cụ lý thuyết hiện có để đối phó với sự không chắc chắn thông tin bao gồm lý thuyết xác suất [3], lý thuyết tập mờ [4], lý thuyết bằng chứng Dempster–Shafer [5,6], lý thuyết entropy thông tin [7,8], v.v. Lý thuyết bằng chứng Dempster–Shafer được sử dụng rộng rãi và công cụ điển hình để xử lý thông tin không chắc chắn và hợp nhất dữ liệu, và nó được sử dụng rộng rãi trong nhiều các lĩnh vực, bao gồm lý do không chắc chắn [9,10], xác định mục tiêu [11], thiết kế bộ điều khiển [12,13], an toàn sản xuất công nghiệp [14], phân loại [15,16], v.v. [17]. Đồng thời, Tham khảo [18]đề xuất quy tắc kết hợp Mô hình niềm tin có thể chuyển đổi (Transferable Belief Model - TBM), đây là phiên bản không chuẩn hóa của quy tắc kết hợp Dempster (DCR). Cả hai quy tắc kết hợp đều có tính chất giao hoán và liên kết, và cả hai đều cho rằng các mục bằng chứng được kết hợp là khác biệt [19]. Tuy nhiên, khi dữ liệu bằng chứng được thu thập có tính xung đột cao, DCR thường không tạo ra kết quả trực quan. Để giải quyết vấn đề này, các học giả đã đề xuất nhiều giải pháp, trong đó đề xuất hàm niềm tin cơ sở [20] dựa trên quy mô khung nhận thức (FOD). Phương pháp này có thể loại bỏ xung đột cao trong bằng chứng bằng cách sửa đổi chỉ định xác suất cơ bản (BPA) trong FOD đầy đủ, do đó tạo ra kết quả trực quan. Nó phù hợp cho các hệ thống quân sự và thời gian thực khác cập nhật hệ thống [20]. Đồng thời, chức năng niềm tin cơ sở đã được sử dụng rộng rãi và mở rộng kể từ khi nó được đề xuất, chẳng hạn như trong [21,22]. Tuy nhiên, chức năng niềm tin cơ sở có một số hạn chế. Nó không xem xét thông tin không chắc chắn do FOD không đầy đủ gây ra, vì vậy nó chỉ có thể được sử dụng trong FOD đầy đủ.

Quy tắc kết hợp tổng quát (The generalized combination rule-GCR) trong FOD không đầy đủ có cùng một vấn đề như DCR cổ điển. Khi đối mặt với bằng chứng mâu thuẫn cao, nó thường tạo ra kết quả kết hợp điều đó trái ngược với trực giác. Hơn nữa, với giả định FOD không đầy đủ, các nguồn thông tin không chắc chắn phức tạp hơn. Trong số đó, thông tin không chắc chắn được đại diện bởi hàm khối lượng khác không của tập hợp rỗng và tính không đầy đủ có thể có của FOD [23] bị bỏ qua bởi hàm niềm tin cơ sở. Để giải quyết các vấn đề trên, bài báo này mở rộng cơ sở niềm tin chức năng và đề xuất một phương pháp để sửa đổi BPA trong FOD không đầy đủ. Phương pháp này không chỉ kế thừa các đặc điểm ban đầu của chức năng niềm tin cơ sở, mà còn tính đến giá trị của hàm khối lượng tập hợp rỗng khác không, cho phép sửa đổi BPA ở trạng thái mở FOD không cạn kiệt. Phương pháp này không chỉ có thể được áp dụng cho FOD không đầy đủ, mà còn có thể được giảm xuống chức năng niềm tin cơ sở trong FOD đầy đủ. Hơn nữa, chúng tôi còn thấy rằng entropy tăng đáng kể sau khi BPA được sửa đổi bằng phương pháp đề xuất. Niềm tin entropy được sử dụng để đo lường sự không chắc chắn của thông tin, có thể hiển thị xung đột một cách trực quan hơn. Sự gia tăng về giá trị của niềm tin entropy là hệ quả của xung đột. Bài báo này cũng đề xuất một dữ liệu xung đột phương pháp quản lý dựa trên hàm niềm tin cơ sở mở rộng (EBBF) và xác minh tính khả thi và hiệu quả của phương pháp đề xuất thông qua việc phân tích một số ví dụ. Các bước xung đột phương pháp quản lý dữ liệu như sau: Đầu tiên, chúng tôi tính toán giá trị của EBBF, sau đó sửa đổi BPA tương ứng với phương pháp sửa đổi được đề xuất và cuối cùng sử dụng DCR hoặc GCR cho dữ liệu tổng hợp để thu được kết quả hợp lý.

Phần còn lại của bài viết này là như sau:

+ Phần 2 giới thiệu các công việc liên quan đến nền kiến thức.

+ Phần 3 giới thiệu kiến thức sơ bộ.

+ Phần 4 đề xuất EBBF và một số các đặc tính của nó.

+ Phần 5 giới thiệu phương pháp quản lý xung đột dựa trên EBBF và đưa ra một số ví dụ và ứng dụng để kiểm chứng tính hiệu quả của phương pháp. Hơn nữa, trong phần 5, sự khác biệt giữa hai quy tắc kết hợp được đề cập ở trên cũng được so sánh và thảo luận.

+ Phần 6 rút ra kết luận của bài viết này.

1. **Công việc liên quan**

Trong lĩnh vực hợp nhất thông tin và quản lý dữ liệu xung đột, các học giả đã áp dụng nhiều phương pháp quản lý dữ liệu xung đột [24]. Hạng mục đầu tiên là xây dựng lại các quy tắc kết hợp. Nhiều nhà nghiên cứu cố gắng xây dựng lại quy tắc kết hợp để xử lý các dữ liệu có tính xung đột cao [23,25,26]. Trong số đó, Yager chỉ ra rằng cần phải loại bỏ yếu tố chuẩn hóa và đặt nó vào miền chưa biết để có được một quyết định hợp lý [27]. Sau đó, dựa trên ý tưởng của Yager, Dubois và Prade đã đề xuất một toán tử kết hợp cụ thể hơn [28]. Lefevre et al, phát triển một khuôn khổ chung để thống nhất một số quy tắc kết hợp cổ điển, có tính đến việc sử dụng đào tạo bộ và tối thiểu hóa hệ số trọng số của tiêu chí lỗi [29]. Su et al. đề xuất một quy tắc mới để kết hợp các phần của bằng chứng từ quan điểm về sự phụ thuộc và độc lập của bằng chứng [30].

Loại thứ hai là tiền xử lý chức năng khối lượng của bằng chứng [31,32]. Quá trình tiền xử lý BPA có thể loại bỏ hiệu quả các hiện tượng hoàn toàn mâu thuẫn giữa bằng chứng, có thể tránh được các kết quả hợp nhất không trực quan do xung đột bằng chứng gây ra [20]. Việc phủ định BPA cũng được đề xuất để giải quyết thông tin không chắc chắn trong bằng chứng [33,34]. Chiến lược trong bài báo này là xử lý trước BPA. Hiện nay, các phương pháp nghiên cứu chính của BPA tiền xử lý như sau.

* 1. Phương pháp tiền xử lý BPA dựa trên tập mờ

Trong trường hợp sử dụng phương pháp tạo tri thức trước cho các mẫu, Xu et al. được tạo ra một BPA lồng nhau sử dụng hàm mật độ xác suất [35]. Loại BPA này vượt trội hơn BPA với những người độc thân ở một mức độ nào đó vì nó ít gây ra xung đột hơn. Tuy nhiên, trên thực tế, nền công nghiệp hiện đại và công nghệ đang phát triển nhanh chóng. Do đó, rất có thể có một số điều chưa biết các loại phần tử bên ngoài FOD đã được thiết lập trước đó [23,36]. Để giải quyết loại vấn đề, Zhang và Deng đã đề xuất một phương pháp xác định điểm biên dựa trên tam giác số mờ [37]. Bằng cách xác định giá trị trung bình, độ lệch chuẩn, giá trị cực trị và tam giác chức năng thành viên của từng thuộc tính và sử dụng giao điểm của mẫu thử nghiệm và mẫu ở trên mô hình, một hàm BPA lồng nhau đã được xây dựng trong [37], có thể gán giá trị cho các tập hợp trống. Phương pháp này phù hợp với cả FOD đầy đủ và FOD không đầy đủ, yêu cầu ít dữ liệu trước hơn và được thúc đẩy bởi dữ liệu, vì vậy nó có thể làm giảm sự không chắc chắn chủ quan. Ngoài ra, nhiều học giả đã đề xuất phương pháp đo độ không đảm bảo dựa trên tập mờ, chẳng hạn như [38–41].

* 1. Phương pháp tiền xử lý BPA với Entropy niềm tin

Trong lĩnh vực tổng hợp thông tin, nhiều học giả đã đề xuất niềm tin entropy để đo lường thông tin không chắc chắn [42–46]. Có thể sử dụng hệ số trọng số được tính toán bởi niềm tin entropy để sửa đổi dữ liệu xung đột [47]. Shannon entropy [48] được áp dụng để đo lường thông tin không chắc chắn trong một khuôn khổ xác suất, và đã được công nhận rộng rãi và mở rộng cho nhiều lĩnh vực, chẳng hạn như entropy mạng trong các mạng phức tạp [49] và phân tích khuếch đại gen trong lĩnh vực tin sinh học [50]. Tuy nhiên, Shannon entropy không thể được áp dụng trực tiếp cho phép đo thông tin không chắc chắn của hàm khối lượng trong lý thuyết bằng chứng. Để giải quyết vấn đề này vấn đề, nhiều biện pháp không chắc chắn trong lý thuyết bằng chứng đã được đề xuất, bao gồm cả sự mơ hồ đo lường [51], Deng entropy [52], v.v. [53,54]. Deng entropy [52] xem xét sự không chắc chắn thông tin được mang bởi hàm khối lượng. Deng entropy không chỉ có thể suy biến thành Shannon entropy trong những điều kiện nhất định mà còn đưa ra phép đo hợp lý trong nhiều môi trường phức tạp [48]. Hơn nữa, Deng entropy đã được sử dụng trong nhiều ứng dụng thực tế, chẳng hạn như chẩn đoán lỗi [55], ra quyết định [56] và hợp nhất dữ liệu cảm biến [34,57]. Tuy nhiên, Deng entropy không tính đến tính đến quy mô của FOD, đây cũng là một nguồn thông tin không chắc chắn quan trọng. Sự thiếu thông tin như vậy có thể dẫn đến giảm hiệu quả xử lý thông tin và thậm chí không có khả năng đối phó với một số thông tin không chắc chắn một cách hiệu quả. Dựa trên kích thước của FOD, Tang et al. đã đề xuất Deng entropy có trọng số [58], làm mất tính nhất quán xác suất mà Deng entropy hài lòng, nhưng mô hình hóa thông tin không chắc chắn hơn trong cơ thể bằng chứng. Sau khi BPA được sửa đổi bằng phương pháp này, sự mất mát thông tin được giảm thiểu một cách hiệu quả và kết quả hợp nhất hợp lý hơn có thể được thu được. Tuy nhiên, Deng entropy không xác minh các thuộc tính cần thiết cho loại phép đo này và trình bày một số hành vi không mong muốn [59] trong một số trường hợp. Mặc dù một số học giả đã đề xuất một biến đổi Deng entropy [58,60], công trình trong [61] đã chứng minh rằng những biến đổi này vẫn không thể đáp ứng hầu hết các tính chất toán học cần thiết và chúng đã trình bày hầu hết các khiếm khuyết về hành vi tồn tại trong Deng entropy. Do đó, Deng entropy và sửa đổi tương ứng phải là thận trọng trong các ứng dụng thực tế.

* 1. Phương pháp tiền xử lý BPA sử dụng hàm niềm tin cơ sở

Đối với phương pháp xây dựng lại luật kết hợp, Haenni đã đề xuất trong bài báo [62] rằng khi có quá nhiều bằng chứng, việc gán cho từng yếu tố trọng lượng là không thực tế và khi có nhiều tập hợp con trong FOD, số tiền tính toán tăng theo cấp số nhân. Để sửa đổi chức năng khối lượng, nó cần ghi lại lượng dữ liệu và tính toán mức độ tương tự hoặc tương quan của dữ liệu, điều này làm tăng thời gian tính toán nên khó thực hiện trong trường hợp yêu cầu thời gian thực cao. trong tầm nhìn của giới hạn trên, Tài liệu tham khảo [20] đề xuất một hàm niềm tin cơ sở mới. Phương pháp này duy trì các đặc điểm tốt của DCR và độ phức tạp tính toán thấp. Hơn nữa, phương pháp này có thể loại bỏ sự mâu thuẫn hoàn toàn giữa các bằng chứng. Dựa trên hàm niềm tin cơ sở, nhiều phương pháp đã được đề xuất bởi các học giả khác, chẳng hạn như của [22,63], có thể giải quyết vấn đề mà DCR không thể thu được kết quả trực quan khi áp dụng cho dữ liệu có tính xung đột cao.Tuy nhiên, các phương pháp trên không phù hợp với FOD không triệt để. Phương pháp đề xuất trong bài báo này là một phần mở rộng của hàm niềm tin cơ sở để làm cho nó có thể áp dụng cho FOD không đầy đủ.

1. **Sơ bộ**

Một số sơ bộ được giới thiệu ngắn gọn trong phần này, bao gồm Dempster–Shafer (D-S) lý thuyết bằng chứng [5,6], hàm niềm tin cơ sở [20], DCR [5,6], quy tắc kết hợp tổng quát [23] v..v...

* 1. Lý thuyết bằng chứng Dempster–Shafer cổ điển

**Định nghĩa 1:** Khung phân biệt Ω được định nghĩa là một tập hữu hạn không rỗng chứa N loại trừ lẫn nhau sự kiện và biểu hiện cụ thể của nó như sau:

Ω = { Θ1, Θ2, ..., Θi , ..., ΘN } (1)

**Định nghĩa 2:** Cho Ω, phép gán xác suất cơ bản (BPA) (hay hàm khối lượng) là ánh xạ m: 2Ω → [0, 1], đáp ứng các thuộc tính sau:

m(∅) = 0, = 1 (2)

Nếu m(A) > 0 thì tập con A được gọi là phần tử tiêu điểm và m(A) > 0 là giá trị hàm khối lượng của mệnh đề tập con A

**Định nghĩa 3:** Trong lý thuyết bằng chứng D-S, DCR có thể hợp nhất hai hàm khối độc lập, m1 và m2:

m(A) = (m1 ⊕ m2)(A) = (3)

trong đó k là hệ số chuẩn hóa được định nghĩa như sau:

k = (4)

Điều đáng chú ý là các định nghĩa cổ điển của lý thuyết bằng chứng D-S được định nghĩa và sử dụng trong FOD đầy đủ

**Định nghĩa 4:** Trong giả thuyết FOD không đầy đủ, DCR được mở rộng bởi Deng trong [52]. Giao của tập rỗng và tập rỗng vẫn là tập rỗng thỏa mãn điều kiện ∅1 ∩ ∅2 = ∅. Cho hai BPAs (m1 và m2), quy tắc kết hợp tổng quát (GCR) được định nghĩa như sau:

m (A) = (5)

K = , (6)

m(∅) = m1(∅) · m2(∅), (7)

m(∅) = 1 khi và chỉ khi K = 1 (8)

* 1. Bình thường hóa và Quy tắc kết hợp TBM

**Định nghĩa 5:** Một BPA m được cho là không chuẩn nếu ∅ là một tập hợp tiêu cự (m(∅) 6= 0), có thể được biến đổi bởi phép toán chuẩn hóa được định nghĩa như sau [19]:

m(A) = k ∗ m(A) khi và chỉ khi A # ∅ (9)

thì m(A) = 0 (10)

Cho tất cả A ⊆ Ω, với k = (1 − m(∅))−1 .

**Định nghĩa 6:** Quy tắc kết hợp TBM được ghi chú với ?; giả sử m1 và m2 là hai BPA, và để m1 và m2 là kết quả của sự kết hợp bởi , như sau [19]:

m12(A) = (11)

Cho tất cả A ⊆ Ω

* 1. Hàm niềm tin cơ bản

**Định nghĩa 7:** Là một công cụ quản lý xung đột dữ liệu, hàm niềm tin cơ sở được đề xuất trong [20] và hàm niềm tin cơ sở được định nghĩa là:

mb (Ai) = (12)

trong đó Ai là một tập con tùy ý của Ω, N là kích thước của FOD, và sau đó BPA ban đầu được sửa đổi theo công thức sau:

m‘(Ai) = (13)

Chức năng niềm tin cơ sở có thể loại bỏ hiệu quả xung đột tuyệt đối giữa các dữ liệu và phù hợp với các hệ thống cập nhật thời gian thực lớn, chẳng hạn như hệ thống quân sự.

* 1. Sự mở rộng của Deng Entropy trong giả định FOD không đầy đủ

**Định nghĩa 8:** Phần mở rộng của Deng entropy [52] là một phương pháp đo entropy, kéo dài từ FOD toàn diện đến FOD không đầy đủ, và định nghĩa của nó như sau [64]:

Eebeow = - (14)

trong đó |A| đại diện cho số lượng phần tử chứa trong mệnh đề A, X đại diện cho FOD và | X| đại diện cho tiềm năng của FOD, đại diện cho số lượng các yếu tố được biết đến chính xác trong FOD. [] là một hàm giới hạn trên (hàm CEILING), liên quan đến việc làm tròn biến độc lập, nghĩa là thành một số nguyên không nhỏ hơn biến độc lập; Ví dụ: [0,2] = 1.