**BÁO CÁO TỔNG KẾT**

1. **Mở đầu**
2. **Nội dung và kết quả nghiên cứu**
3. **Chương 1: Cơ sở lý thuyết**
4. Lý thuyết bằng chứng Demster-Shafer

* Lý thuyết D-S (Demster 1967), trình bày bởi Demster và  
  tiếp tục được phát triển bởi Rota (1977), thường được coi là một  
  sự mở rộng của lý thuyết xác suất Bayes. Lý thuyết D-S có thể quản lý để xử lý những sự không chắc chắn gây ra bởi các xác suất trước chưa biết trong khi Bayes chủ quan không thể làm được. Kể từ khi tập trung vào câu hỏi về tính không chắc chắn, nó đã được áp dụng trong nhiều lĩnh vực (Utkin 2009; Kabir và cộng sự 2015; Gong và cộng sự 2018; Yin và Deng 2018). Một số định nghĩa và lý thuyết được đề cập như sau
* Hãy để  là một tập hợp hữu hạn không rỗng của N giá trị có thể mà  
  là loại trừ lẫn nhau.  được gọi là khung phân biệt   
  và được định nghĩa là
  +  = {H1, H2, …. , HN} (1)
* Sức mạnh của tập hợp gồm các mệnh đề 2N của  là  
  đã xác định:
  + = {,{H1},{H2},…,{HN},{H1∪ H2},…,{H1∪ H2∪ … ∪ Hi}, …. } (2)
* Đối với , một bài tập xác suất cơ bản (BPA) (còn được gọi là  
  hàm đại chúng) là một ánh xạ m: → [0, 1], thỏa mãn  
  các thuộc tính sau:



* +   (3)
  +  (4)

1. Nếu A ⊂  và A≠ , sau đó, thì hàm số khối lượng m (A) đại diện cho khả năng của bằng chứng A ủng hộ tuyên bố. Các m(A) lớn hơn là, bằng chứng mạnh hơn ủng hộ giả thuyết A. Bất kỳ tập con A ⊆  chẳng hạn như m(A) > 0 được gọi là một phần tử tiêu cự của m

Đối với , belief function từ BPA m được định nghĩa là

*bel* :  (5)

 (6)

* Số lượng bell (A) có thể được hiểu là thước đo của một người  
  niềm tin rằng giả thuyết A là đúng. Cụ thể, bel()) = 0 và  
  bell() = 1  
  Do nguồn khác nhau, chúng tôi có thể nhận được BPA khác nhau từ cùng một bằng chứng. Demster đề xuất sử dụng tổng trực giao để kết hợp các BPA này, là trao đổi và liên kết. Dựa trên những bằng chứng độc lập, tổng trực giao của các chức năng khối lượng của họ tính toán  
  mức độ tin tưởng vào các quy tắc kết hợp.
  + m = m1 ⨁ m2 ⨁ m3 ⨁ …. ⨁ mn (7)
* trong đó mi là hàm số khối lượng(mass function) của mỗi quy tắc.  
  Giả sử có n BPA được chỉ định bằng m1, m2, …, mn.  
  Các quy tắc kết hợp như sau:

*m*() = 0

 (8)

* trong đó k là một yếu tố bình thường hóa
  +  (9)
* Để cụ thể, giả sử có hai BPA được chỉ định bởi m1 và m2, các yếu tố tiêu điểm là Bi và Ci, sau đó các quy tắc kết hợp được đơn giản hóa như:
  + m(A) = 0 A 

 A =  (10)

* trong đó k đo mức độ xung đột giữa m1 và m2
  +  (11)

* k = 0 có nghĩa là m1 phù hợp với m2. k = 1 có nghĩa là m1 hoàn toàn mâu thuẫn với m2, tức là hai nguồn ủng hộ mạnh mẽ các giả thuyết khác nhau không tương thích.

1. Một phương pháp quản lý xung đột hiệu quả

3. Phương pháp được đề xuất

3.1 Base belief function

* Giả định của chúng ta là trong một thế giới khép kín. Để  là một tập hợp N giá trị có thể độc quyền lẫn nhau. Bộ quyền lực của  là , trong đó số phần tử là 2N. Khi khung phân biệt hoàn tất, m() = 0. Do đó, chúng tôi xác định hàm niềm tin cơ bản mb là

mb(Ai) = (12)

trong đó Ai là mọi tập con trong  ngoại trừ tập rỗng .

* Sau đó, chúng ta có thể sử dụng các phương pháp khác nhau để tạo ra BPA dựa trên các bằng chứng chúng ta có. Giả sử BPA được chỉ định bởi m, chúng tôi sử dụng mb để sửa m bằng cách tính trung bình số học:

m’(Ai) = (13)

* Mục đích của base belief function là cho mỗi tập con trong khung phân biệt một khả năng tương đương trước khi chúng tạo ra BPA. Hãy xem xét có ba quả bóng giống hệt nhau có màu sắc khác nhau và ít nhất một quả nằm trong một cái túi mờ. Nếu chúng ta không có được những manh mối khác, sau đó chúng ta có thể có được bảy loại tình huống có cùng khả năng. Chức năng niềm tin cơ bản dựa trên những suy nghĩ như vậy. BPA đại diện cho mức độ của một nguồn hỗ trợ cho các giả thuyết trong một tình huống tại một thời điểm nhất định. Nhưng trước khi mọi nguồn tin xuất hiện, niềm tin của chúng ta vào mọi tình huống phải bình đẳng. base belief function tương đương với những khả năng ban đầu này.
* Sau đó, khi một nguồn đến, chúng tôi sử dụng base belief function để điều chỉnh BPA cổ điển. Nếu m(ai) >mb(ai), thì m’(ai) >mb(ai), nghĩa là nguồn sắp tới củng cố khả năng ban đầu. Nếu m(Ai) < mb(Ai), thì m’(A) < mb(Ai), có nghĩa là khả năng ban đầu bị suy yếu.
* Ưu điểm lớn nhất của base belief function là nó loại trừ tình huống mà bằng chứng hoàn toàn mâu thuẫn với nhau. Bằng chứng thường xảy ra xung đột cao độ khi một số khối xác suất là 0, và base belief function có thể tránh được tình huống đó một cách hoàn hảo. Ngay cả khi tất cả các nguồn làm suy yếu một giả thuyết, BPA sửa đổi của giả thuyết này sẽ gần bằng 0, nhưng nó sẽ không bao giờ là 0. Phương pháp như vậy phủ định tính tuyệt đối trong thế giới thực và cho mọi BPA một tỷ lệ chấp nhận lỗi. Nói cách khác, sự không chính xác của các nguồn sẽ không có ảnh hưởng quyết định đến kết quả kết hợp cuối cùng, đặc biệt khi chúng ta cần xử lý một lượng lớn dữ liệu. Ngay cả khi các bằng chứng có sẵn ủng hộ mạnh mẽ giả thuyết A và phản đối giả thuyết B bây giờ, kể từ khi chúng tôi không thu thập tất cả các bằng chứng, và chúng tôi có thể không bao giờ thu thập tất cả các bằng chứng trong hầu hết các tình huống trong thực tế, khả năng rằng A là sai và B là đúng
* Không thể tránh khỏi, phương pháp được đề xuất này trao đổi thời gian để có kết quả kết hợp chính xác và trực quan, đòi hỏi một lượng lớn tính phức tạp. Điều kiện có thể giảm tải tính toán một cách thích hợp được đưa vào Mục 3.3

3.2 Các ví dụ số liệu về việc sử dụng base belief function

* Đầu tiên, chúng tôi đưa ra hai ví dụ đơn giản và cực đoan để xác minh tính đúng đắn của ý kiến đề xuất bản.
* Ví dụ 2 Giả sử FOD là  = {a, b} và hai BPA được cung cấp như

m1(a) = 1, m1(b) = 0, m1(a,b) = 0

m2(a) = 1, m2(b) = 0, m2(a,b) = 0

* Đầu tiên chúng tôi nhận được các base belief function theo công thức (12)

mb(a) = mb(b) = mb(a,b) =

* Sau đó, chúng tôi sửa đổi hai BPA dựa trên công thức (13)

mi’(a) = = 0.6667, mi’(b) = mi’(a,b) = = 0.1667

với i = 1,2

* Sử dụng quy tắc tổ hợp của Demster, chúng ta có thể có được kết quả:

m(a) = 0.8571, m(b) = 0.1071, m(a,b) = 0.0357

* Theo kết quả, chúng tôi có thể phân tích rằng {a} được hỗ trợ mạnh mẽ bởi nguồn gốc, phù hợp với thực tế. Tuy nhiên, chúng tôi vẫn thừa nhận rằng {b} có khả năng trở thành sự thật, mặc dù khả năng là rất nhỏ.
* Ví dụ 3 Giả sử FOD là  = {a, b} và hai BPA được cung cấp như

m1(a) = 1, m1(b) = 0, m1(a,b) = 0

m2(a) = 0, m2(b) = 1, m2(a,b) = 0

Hàm base belief function là:

mb(a) = mb(b) = mb(a,b)

* Sau khi chỉnh sửa hai BPA, chúng ta có thể có được sự kết hợp  
  kết quả sử dụng lý thuyết D-S:
  + m(a) = m(b) = 0.4737, m(a,b) = 0.0526
* Ví dụ này chỉ ra rằng {a} và {b} có khả năng tương đương nhau, điều này phản ánh đúng thực tế vì các nguồn được đưa ra ở trên hoàn toàn xung đột
* Trong Mục 2.2, chúng tôi đã đưa ra một ví dụ và một số lý do cho khả năng gây ra xung đột. Trong tiểu mục này, chúng tôi vẫn sử dụng cùng một ví dụ để kiểm tra hiệu quả của base belief function như sau
* Ví dụ 4 (giống như ví dụ 1) Giả sử rằng FOD  
  là  = {a, b, c} và hai BPA được đưa ra là

m1(a) = 0.99, m1(a,b) = 0.01

m2(b) = 0.01, m2(c) = 0.99

chúng ta nhận được base belief function

mb(a) = mb(b) = mb(c) = mb(a,b) = mb(a,c) = mb(b,c) = mb(a,b,c) =

* Sau đó, chúng tôi sửa đổi hai BPA và có được kết quả cuối cùng:

m(a) = m(c) = 0.3957, m(b) = 0.1012, m(a,b) = 0.0337

m(a,c) = m(b,c) = 0.0322, m(a,b,c) = 0.0107

Như thể hiện trong Bảng 1, kết quả kết hợp của phương pháp được đề xuất hợp lý hơn nhiều so với kết quả kết hợp của Demster cổ điển trong Mục 2.2. m(b) = 0.1012 hợp lý hơn m(b) = 1. Mặc dù m(b) lớn hơn nhiều so với chức năng khối lượng ban đầu, nó vẫn nhỏ hơn 1/7, điều này có nghĩa là {b} m (b) không có quá nhiều hỗ trợ 0.3957 là hợp lý vì {a} và {c} nhận được sự hỗ trợ mạnh mẽ từ các nguồn gốc ban đầu.

* Bảng 1 Kết quả của hai  
  quy tắc kết hợp của Ví dụ 4.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | m(a) | m(b) | m(c) | m(a,b) | m(a,c) | m(b,c) | m(a,b,c) |
| Quy tắc Dempster cổ điển | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Quy tắc đề xuất | 0.3957 | 0.1012 | 0.3957 | 0.0337 | 0.0322 | 0.0322 | 0.0107 |

3.3 Điều kiện thích hợp

* Ví dụ 5 Giả sử FOD là  = {a, b, c} và hai BPA được cung cấp như sau

m1(a) = 0.9, m1(a,b,c) = 0.1

m2(c) = 0.9, m2(a,b,c) = 0.1

Kết quả của hai quy tắc kết hợp được trình bày trong Bảng 2.

* Bảng 2 Kết quả của hai quy tắc kết hợp trong Ví dụ 5

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | m(a) | m(b) | m(c) | m(a,b) | m(a,c) | m(b,c) | m(a,b,c) |
| Quy tắc Dempster cổ điển | 0.4737 | 0 | 0.4737 | 0 | 0 | 0 | 0.0526 |
| Quy tắc đề xuất | 0.3756 | 0.0976 | 0.3756 | 0.0413 | 0.0413 | 0.0413 | 0.0271 |

* Ví dụ 6 Giả sử rằng FOD là  = {a, b, c} và hai BPA được cung cấp như

m1(a) = 0.9, m1(b) = 0.05, m1(c) = 0.05

m2(a) = 0.05, m2(b) = 0.05, m2(c) = 0.9

Kết quả của hai quy tắc kết hợp được trình bày trong Bảng 3.Như có thể thấy trong Bảng 2 và 3, khi tất cả các chức năng khối lượng của một tập hợp hai bằng chứng là không bằng 0 hoặc tất cả các chức năng khối lượng của tập hợp hoàn chỉnh của hai bằng chứng là không bằng 0, kết quả của hai quy tắc này có ít khác biệt. Ví dụ 5 và 6 lần lượt hỗ trợ mạnh mẽ trên {a} và {c}. Kết quả của hai quy tắc tổ hợp đều hợp lý, mang lại khả năng bình đẳng cho {a} và {c} nhiều hơn nhiều so với {b}. Để giảm độ phức tạp trong tính toán, không cần phải dùng base belief function khi hai tình huống này xuất hiện.

* Bảng 3 Kết quả của hai quy tắc kết hợp trong Ví dụ 6

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | m(a) | m(b) | m(c) | m(a,b) | m(a,c) | m(b,c) | m(a,b,c) |
| Quy tắc Dempster cổ điển | 0.4865 | 0.0270 | 0.4865 | 0 | 0 | 0 | 0.0526 |
| Quy tắc đề xuất | 0.3876 | 0.1222 | 0.3876 | 0.0308 | 0.0308 | 0.0308 | 0.0103 |

* Ví dụ 7 Giả sử rằng FOD là  = {a, b, c} và haiBPA được cung cấp như

m1(a) = 0.9, m1(a,b,c) = 0.1

m2(a) = 0.05, m2(b) = 0.05, m2(c) = 0.9

Kết quả của hai quy tắc kết hợp được trình bày trong Bảng 4.

* Bảng 4: kết quả của hai quy tắc tổ hợp của ví dụ 7

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | m(a) | m(b) | m(c) | m(a,b) | m(a,c) | m(b,c) | m(a,b,c) |
| Quy tắc Dempster cổ điển | 0.3448 | 0.0345 | 0.6207 | 0 | 0 | 0 | 0 |
| Quy tắc đề xuất | 0.3791 | 0.1110 | 0.3846 | 0.0362 | 0.0362 | 0.0362 | 0.0166 |

* Có thể dễ dàng thấy rằng kết quả từ phương pháp được đề xuất của chúng tôi hợp lý hơn so với quy tắc của Demster cổ điển. Nguồn 1 hỗ trợ mạnh mẽ {a}, trong khi nguồn 2 ít hỗ trợ trên {a} và ngược lại hỗ trợ {c}. Mặc dù có rất ít hỗ trợ trên {b} nói chung được hai nguồn chấp nhận, nhưng kết quả kết hợp cung cấp hỗ trợ tương đương cho {a} và {c} là hợp lý hơn
* Có thể quan sát và thảo luận thêm nhiều điều nữa. Với ví dụ 5 và 6, tập hợp đầy đủ của kết quả kết hợp không phải là 0 ngay cả khi chúng ta sử dụng quy tắc Demster cổ điển. Trong những tình huống như vậy, hai quy tắc tổ hợp có tác động gần như giống nhau. Tuy nhiên, ví dụ 7 tập hợp đầy đủ kết quả kết hợp là 0, và hai kết quả có sự khác biệt rõ rệt. Do đó, chúng ta có thể kết luận rằng trong những tình huống có xung đột, toàn bộ tập hợp có thể được coi là yếu tố làm suy yếu xung đột. Mỗi khi chúng ta tạo ra BPA, nhiều xung đột có thể được tránh nếu các hàm số lượng của tập hợp hoàn chỉnh không phải là 0.

3.4 Quy trình của phương pháp đề xuất sử dụng base belief function

* Bước 1: Tạo base belief function cho dù bằng chứng là một hệ thống ứng dụng cập nhật theo thời gian thực hoặc một tập dữ liệu, mỗi thuộc tính được coi là một nguồn thông tin độc lập. Sử dụng các loại khác nhau của phương pháp tạp BPA của từng thuộc tính là tầm quan trọng đặc biệt là trong hệ thống quân sự. Điều duy nhất chúng tôi cần quan tâm là liệu bằng chứng sắp tới có thuộc về một trong hai điều kiện được đề cập ở trên. Nếu vậy, hãy liên kết trực tiếp với kết quả được tạo ra; nếu không, sử dụng cơ sở chức năng niềm tin để sửa đổi BPA trước khi kết hợp.
* Bước 2: Đánh giá điều kiện
  + Điều kiện 1: Tất cả các hàm khối lượng của các bộ đơn lẻ hai BPA là khác 0.
  + Điều kiện 2: Tât cả các chức năng khối lượng của bộ hoàn chỉnh trong số hai BPA là khác 0.
  + Điều kiện 3: Không thuộc điều kiện 1 và điều kiện 2.
  + chúng ta phải tìm ra điều kiện nào mà hai BPA này thuộc. Mục đích của bước này là để giảm thiểu số tiền tính toán.
* Bước 3: sử dụng quy tắc kết hợp thích hợp Theo Ví dụ 5 và 6 đã nói ở trên, nếu hai BPA thuộc Điều kiện 1 hoặc Điều kiện 2, chỉ có quy tắc của clas-sical Dempster được yêu cầu để kết hợp hai Nguồn. Tuy nhiên, nếu hai BPA thuộc về Điều kiện 3, chúng ta cần sử dụng chức năng niềm tin cơ bản và tạo ra một BPA được sửa đổi từ mỗi nguồn độc lập của thông tin. BPA sửa đổi có thể được sử dụng vào Quy tắc kết hợp của Dempster. Cuối cùng, một kết quả kết hợp thu được và số lượng của tất cả các BPA được giảm một. Lặp lại các bước 1–3, chúng ta có thể đạt được kết quả cuối cùng khi chỉ có một BPA Bên trái. Sự khác biệt duy nhất giữa thủ tục đầu tiên và lặp lại thủ tục là trong việc lặp lại proce-dure, một trong hai BPAs là kết quả kết hợp từ thủ tục cuối cùng. Hình 1 cho thấy lưu đồ của phương pháp đề xuất.

Hình 1: lưu đồ



* Thủ tục như vậy bảo toàn những đặc tính quý giá của quy tắc của Demster cổ điển như tính kết hợp và tính giao hoán. Một ưu điểm khác của phương pháp được đề xuất là có thể đạt được sự hợp nhất tuần tự. Chúng ta có thể kết hợp các bằng chứng theo thứ tự xuất hiện của chúng thay vì xem xét thứ tự kết hợp của chúng, điều quan trọng là có thể đạt được sự hợp nhất theo tuần tự. Chúng ta có thể kết hợp các bằng chứng theo thứ tự xuất hiện của họ vì xem xét thứ tự kết hợp của họ, điều này rất quan trọng.

4. Thí nghiệm

* Hai thí nghiệm phân loại sử dụng bộ dữ liệu thực được thực hiện để đánh giá hiệu quả của base belief function trong phần này. Các bộ dữ liệu thật là từ UCI kho lưu trữ máy học (Bache và Richman 2013).

4.1 Thí nghiệm 1(Phân loại bộ dữ liệu Iris)

* Bảng 5: Số fuzzy tam giác của 4 thuộc tính

|  |  |  |  |
| --- | --- | --- | --- |
|  | a | b | c |
| SL | (4.30, 5.03, 5.80) | (4.90, 6.01, 7.10) | (5.40, 6.68, 7.90) |
| SW | (3.00, 3.42, 4.40) | (2.20, 2.77, 3.20) | (2.50, 2.97, 3.80) |
| PL | (1.20, 1.46, 1.70) | (2.20, 4.26, 5.00) | (4.80, 5.55, 6.90) |
| PW | (0.10, 0.23, 0.40) | (1.00, 1.30, 1.60) | (1.50, 2.03, 2.50) |

* Có ba loài (Setosa (a), Vsicolor (b), Virginica (c) trong bộ dữ liệu Iris với bốn thuộc tính và từng loài chứa 50 trường hợp. Chúng tôi chọn ngẫu nhiên 40 trường hợp từ mỗi loài và tạo ra các số fuzzy tam giác (Klir và Yuan 1996) của bốn thuộc tính được thể hiện trong Bảng 5.  
   Mười trường hợp còn lại được coi là bộ kiểm tra. Chúng tôi chọn ngẫu nhiên một cá thể từ loài Setosa của bộ thử nghiệm và tạo BPA của nó. Kết quả được hiển thị trong Bảng 6. Bốn thuộc tính của trường hợp này là (5.3, 3.5, 1.3, 0.2).
* Bảng 6: BPAs của 4 thuộc tính

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | m(a) | m(b) | m(c) | m(a,b) | m(a,c) | m(b,c) | m(a,b,c) |
| SL | 0.3337 | 0.3165 | 0.2816 | 0.0307 | 0.0052 | 0.0272 | 0.0052 |
| SW | 0.3164 | 0.2501 | 0.2732 | 0.0304 | 0.0481 | 0.0515 | 0.0304 |
| PL | 0.6699 | 0.3258 | 0 | 0 | 0 | 0.0043 | 0 |
| PW | 0.6996 | 0.2778 | 0 | 0 | 0 | 0.0226 | 0 |

Sau đó chúng tôi sử dụng phương pháp được đề xuất để có được kết quả cuối cùng. Cụ thể hơn, hàm số khối lượng của thuộc tính PL và PW cần phải được sửa đổi bằng base belief function, vì chúng không thỏa mãn một trong hai điều kiện cụ thể. Bảng 7 cho thấy BPA đã được sửa đổi.

* Bảng 7: BPAs của 4 thuộc tính sử dụng hàm đức tin cơ bản

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | m(a) | m(b) | m(c) | m(a,b) | m(a,c) | m(b,c) | m(a,b,c) |
| SL | 0.3337 | 0.3165 | 0.2816 | 0.0307 | 0.0052 | 0.0272 | 0.0052 |
| SW | 0.3164 | 0.2501 | 0.2732 | 0.0304 | 0.0481 | 0.0515 | 0.0304 |
| PL | 0.4064 | 0.2343 | 0.0714 | 0.0714 | 0.0714 | 0.0736 | 0.0714 |
| PW | 0.4213 | 0.2103 | 0.0714 | 0.0714 | 0.0714 | 0.0827 | 0.0714 |

Kết quả theo các quy tắc kết hợp khác nhau được trình bày trong Bảng 8.

* Bảng 8: Kết quả của các quy tắc kết hợp khác nhau của thí nghiệm Iris

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Quy tắc Yager của Demster cổ điển (Yager 1987) Luật lệ của Deng và cộng sự (Yong và cộng sự). 2004) Quy tắc được đề xuất | m(a) | m(b) | m(c) | m() |
| 0.8457 | 0.1543 | 0 | 0 |
| 0.5337 | 0.1484 | 0 | 0.3180 |
| 0.8533 | 0.1361 | 0.0119 | 0 |
| 0.6232 | 0.2671 | 0.1083 | 0 |

Như có thể thấy trong Bảng 8, tất cả các quy tắc này có thể nhận ra rằng trường hợp thử nghiệm có thể là Setosa, phù hợp với tình hình thực tế. Mặc dù kết luận của quy tắc Demster cổ điển có vẻ rõ ràng hơn, nó tồn tại rủi ro của xung đột vì m (c) = 0. Nếu có nhiều thuộc tính hơn và m(a) = 0, kết quả sẽ không hợp lý. Vì mục đích chính xác và bảo mật, phương pháp được đề xuất và quy tắc của Deng và cộng sự tốt hơn nhiều. Trong tiểu mục tiếp theo, ưu điểm của base belief function sẽ rõ ràng hơn nhiều.

4.2 Thí nghiệm 2(bộ dữ liệu rượu)

* Trong thử nghiệm này, chúng tôi sử dụng phân loại tập dữ liệu rượu vang để chứng minh độ rộng ứng dụng của base belief function. Bộ dữ liệu rượu bao gồm ba loại rượu khác nhau với 13 thuộc tính. Chúng tôi chọn ngẫu nhiên 45 yếu tố từ mỗi loại rượu làm mẫu huấn luyện và các yếu tố còn lại trong mỗi loại làm mẫu thử nghiệm. Sau đó, chúng tôi chọn một ví dụ, một ví dụ cực đoan, từ loại a và tạo BPA của nó. 13 thuộc tính của phiên bản thử nghiệm này là (13,24, 3,98, 2,29, 17,5, 103, 2,64, 2,63, 0,32, 1,66, 4,36, 0,82, 3, 680). Quá trình thử nghiệm được đề cập trong Phần. 4.1. Để đơn giản, chỉ các BPA thu được và kết quả kết hợp cuối cùng được trình bày trong Bảng 9 và 10 tương ứng
* Bảng 9: BPAs của 13 thuộc tính

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | m(a) | m(b) | m(c) | m(a, b) | m(a, c) | m(b, c) | m(a, b, c) |
| alcohol | 0.1304 | 0.1160 | 0.2752 | 0.1160 | 0.1304 | 0.1160 | 0.1160 |
| malic acid | 0.3082 | 0.1681 | 0.0889 | 0.1681 | 0.0889 | 0.0889 | 0.0889 |
| ash | 0.1413 | 0.1994 | 0.1295 | 0.1413 | 0.1295 | 0.1295 | 0.1295 |
| alcalinity of ash | 0.3966 | 0.3017 | 0 | 0.3017 | 0 | 0 | 0 |
| magnesium | 0.2252 | 0.2372 | 0.2252 | 0.2252 | 0.0781 | 0.0781 | 0.0781 |
| total phenols | 0.2599 | 0.2642 | 0.0540 | 0.2599 | 0.0540 | 0.0540 | 0.0540 |
| flavanoids | 0.2885 | 0.4230 | 0 | 0.2885 | 0 | 0 | 0 |
| nonflavanoid phenols | 0.1840 | 0.1776 | 0.1152 | 0.1776 | 0.1152 | 0.1152 | 0.1152 |
| proanthocyanins | 0.1301 | 0.2030 | 0.1383 | 0.1301 | 0.1301 | 0.1383 | 0.1301 |
| color intensity | 0.2123 | 0.2849 | 0.0726 | 0.2123 | 0.0726 | 0.0726 | 0.0726 |
| hue | 0 | 0.2921 | 0.4158 | 0 | 0 | 0.2921 | 0 |
| OD280∕OD315 of diluted wines | 0.3314 | 0.3372 | 0 | 0.3314 | 0 | 0 | 0 |
| proline | 0 | 0.3106 | 0.3788 | 0 | 0 | 0.3106 | 0 |

* Bảng 10: Kết quả khác nhau  
  quy tắc pha rượu  
  thí nghiệm

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Quy tắc Yager của Demster cổ điển (Yager 1987) Luật lệ của Deng và cộng sự (Yong và cộng sự). 2004) Quy tắc được đề xuất | m(a) | m(b) | m(c) | m(a,b) | m(a,c) | m(b,c) | m() |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0.4371 | 0.1237 | 0 | 0 | 0.1014 | 0.3370 |
| 0.1914 | 0.7820 | 0.0010 | 0 | 0 | 0 | 0 |
| 0.4997 | 0.4569 | 0.0434 | 0 | 0 | 0 | 0 |

* Như thể hiện trong Bảng 10, kết quả kết hợp của quy tắc Dempster cổ điển minh họa rằng cá thể thử nghiệm hoàn toàn thuộc về loại b. Kết quả của phương pháp Yager gợi ý rằng có khả năng đối tượng thử nghiệm là loại b và chúng ta vẫn còn ít kiến thức. Kết quả của quy tắc Deng và cộng sự cũng gợi ý rằng trường hợp thử nghiệm có khả năng là biến thể b. Chỉ có phương pháp được đề xuất của chúng tôi minh họa rằng có một khả năng nhỏ là đối tượng thử nghiệm thuộc loại a hơn b, điều này phù hợp với tình hình thực tế. Rõ ràng, phương pháp được đề xuất tốt hơn nhiều so với các quy tắc kết hợp khác trong thí nghiệm này. Lý do cho sự khác biệt lớn của các kết quả kết hợp này là do tồn tại nhiều xung đột trong việc tạo ra BPA, điều này xác minh việc quản lý xung đột base belief function. Từ thử nghiệm này, chúng tôi có thể kết luận rằng hiệu suất của phương pháp được đề xuất của chúng tôi tốt hơn phương pháp của Deng và cộng sự trong tình huống dữ liệu đại chúng.

5. Kết luận

* Khi kết hợp những niềm tin trái ngược nhau bằng cách sử dụng các quy tắc của Dempster, kết quả thường phản trực giác. Trong số tất cả các phương pháp thay thế giải quyết vấn đề, chúng sẽ thay đổi quy tắc kết hợp hoặc sửa đổi mô hình dữ liệu. Trong bài báo này, phương pháp được đề xuất xác định hàm niềm tin cơ sở dựa trên khung phân biệt, làm cho các hàm khối lượng của mọi tập con khác không để tránh vấn đề về quy tắc Dempster cổ điển, đây là một giải pháp cơ bản để quản lý xung đột. Hơn nữa, miễn là các hàm khối lượng của các tập hợp đơn lẻ hoặc tập hợp đầy đủ khác không, thì kết quả tổ hợp sẽ logic. Phương pháp như vậy bảo tồn một số tính chất mong muốn như tính giao hoán và tính kết hợp.
* Tuy nhiên, phương pháp đề xuất có một số hạn chế. Base belief function chỉ có thể được sử dụng trong một thế giới khép kín. Trong một thế giới mở, có nhiều khả năng gây ra xung đột, hoặc khung nhận thức không đầy đủ hoặc các cảm biến không đáng tin cậy. Sẽ không khôn ngoan nếu giải quyết xung đột theo một cách chung chung. Ngoài ra, hiệu quả của phương pháp đề xuất là không rõ ràng khi tập dữ liệu khá nhỏ. Vì nó cung cấp khả năng chịu lỗi cho mọi cảm biến, nên kết quả cuối cùng chỉ hiển nhiên là chính xác nếu các tập dữ liệu lớn xuất hiện.
* Phương pháp đề xuất của chúng tôi có thể được sử dụng trong việc ra quyết định hoặc nhận dạng mẫu trong trường hợp không chắc chắn mạnh, đặc biệt phù hợp với tình huống dữ liệu lớn như hệ thống đa tác nhân hoặc hệ thống quân sự. Trong những tình huống này, bằng chứng mâu thuẫn cao thường xảy ra do nhiều nguồn thông tin và các phép đo khác thường, khiến cho quy tắc Dempster cổ điển là không thể. Sử dụng hàm niềm tin cơ sở để sửa đổi BPA, mặc dù kết quả tổ hợp có thể không rõ ràng trong các tập dữ liệu nhỏ do mâu thuẫn, nhưng nó sẽ dần dần đạt được câu trả lời chính xác do các tập dữ liệu lớn. Do đó, chúng ta có thể kết hợp các bằng chứng theo thời gian đến của chúng mà không cần xem xét thứ tự kết hợp, điều này hoàn toàn phù hợp với hệ thống ứng dụng được cập nhật theo thời gian thực cao.

1. 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
2. **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:

 (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:

,  (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:

 (3)

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

 (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:

 , (5)

, (6)

, (7)

 khi và chỉ khi . (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]:

 khi và chỉ khi  (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]:

 (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:

, (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:

 (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]:

 (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.

1. **Hàm niềm tin cơ sở mở rộng**

Trong phần này, chúng tôi đề xuất một hàm phân phối niềm tin trong FOD không đầy đủ dựa trên hàm niềm tin cơ sở, sử dụng giá trị của hằng số tự nhiên e và hàm khối lượng tập rỗng như một yếu tố đặc trưng mà hàm này được sử dụng để sửa đổi BPA. Trong khi đó, phương pháp này cũng được áp dụng cho FOD toàn diện

Người ta cho rằng FOD A chứa N phần tử loại trừ lẫn nhau; khi đó, A có 2 tập con N và trong FOD không đầy đủ, giá trị của hàm khối lượng tập rỗng không phải là 0. Dựa trên các giả định trên, hàm gán niềm tin cơ bản mở rộng được định nghĩa như sau:

, (15)

trong đó Ri đại diện cho một tập con tùy ý của FOD, N là kích thước của FOD và m(∅) là giá trị hàm khối lượng của tập rỗng.

Sử dụng hàm phân phối niềm tin cơ sở mở rộng neb (Ri), BPA của mỗi bằng chứng được sửa đổi và phương pháp sửa đổi được định nghĩa như sau:

, (16)

trong đó m(Ri) là giá trị của BPA của tất cả các bằng chứng. 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.

Mục đích của phép tính neb(Ri) là gán xác suất bằng nhau cho mỗi tập con của FOD trước khi BPA được tạo ra. Trước khi mọi nguồn bằng chứng xuất hiện, niềm tin của chúng ta vào mọi tình huống phải bình đẳng. neb (Ri) tương đương với những khả năng ban đầu này, có thể loại bỏ hiện tượng mâu thuẫn hoàn toàn giữa các bằng chứng. Phương pháp này phủ nhận tính tuyệt đối của bằng chứng, nghĩa là, ngay cả khi tất cả các bằng chứng hiện có đều chống lại một tập hợp con A, nhưng chúng tôi chưa tìm thấy tất cả các bằng chứng, vẫn có khả năng A là đúng. Đồng thời, giá trị khối lượng nonzero của tập rỗng được thêm vào neb (Ri) và quá trình hợp nhất làm hệ số đặc trưng. Nghĩa là, trong FOD không đầy đủ, sự không chắc chắn thông tin gây ra bởi hàm khối lượng tập rỗng nonzero được tính đến trong công thức để phương pháp có thể được sử dụng trong FOD không đầy đủ và BPA sửa đổi vẫn đáp ứng tính chất cơ bản là tổng các giá trị hàm khối lượng của tất cả các tập con là 1.

1. **Áp dụng hàm niềm tin cơ sở mở rộng**

Trong phần này, chúng tôi đề xuất một phương pháp quản lý dữ liệu xung đột dựa trên EBBF và đưa ra một số ví dụ và ứng dụng liên quan để xác minh tính hợp lý và hiệu quả của nó.

* 1. Phương pháp quản lý dữ liệu xung đột dựa trên EBBF

Phần này đề xuất một phương pháp quản lý dữ liệu xung đột dựa trên EBBF để xác minh khả năng ứng dụng và hiệu quả của EBBF trong lĩnh vực hợp nhất thông tin. Hình 1 thiết kế khuôn khổ của phương pháp quản lý dữ liệu xung đột dựa trên EBBF và các bước chi tiết như sau:

Bước 1: Trong FOD không đầy đủ, có rất nhiều thông tin không chắc chắn trong các ứng dụng thực tế. Để mô hình hóa thông tin không chắc chắn một cách có hệ thống, trong khuôn khổ của lý thuyết bằng chứng Dempster-Shafer, bước đầu tiên là sử dụng BPA để mô hình hóa thông tin không chắc chắn.

Bước 2: Tính toán giá trị của EBBF: Khi dữ liệu có nhiều xung đột, cần có một phương pháp hợp lý và hiệu quả để xử lý trước BPA trước khi xử lý dữ liệu tiếp theo. Phương pháp này sử dụng EBBF để sửa đổi BPA. Công thức tính tương ứng của hàm niềm tin cơ sở mở rộng (BBF) như sau:

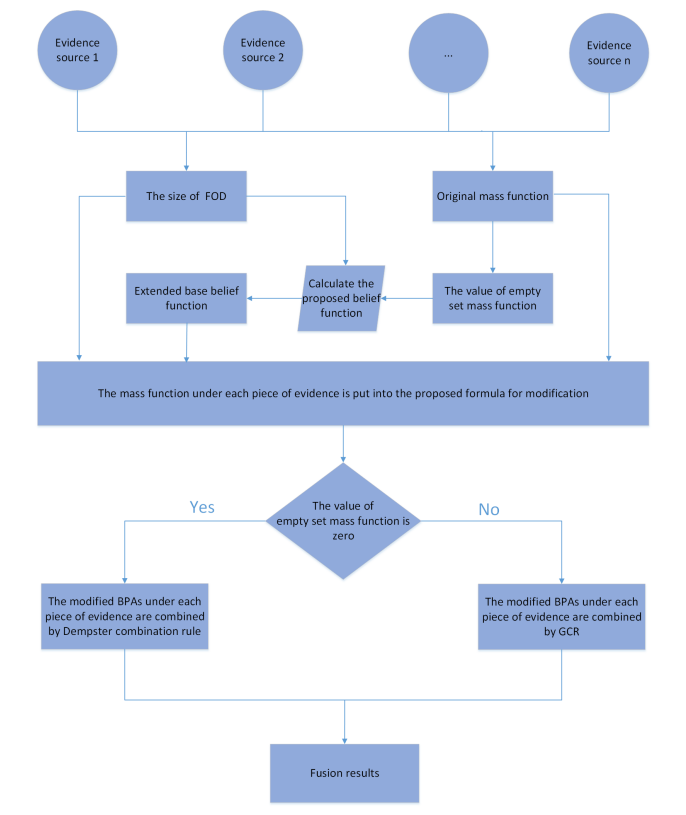
 (17)

Bước 3: Theo giá trị EBBF được tính toán, BPA của mỗi nhóm được sửa đổi và công thức sửa đổi như sau:

 (18)

Bước 4: Theo đó hàm khối lượng tập rỗng là 0, DCR hay GCR được chọn để hợp nhất dữ liệu. Theo DCR (n − 1), kết quả kết hợp của mỗi mệnh đề có thể thu được như sau:

.(19)



Hình 1. Khung phương pháp quản lý dữ liệu xung đột dựa trên hàm niềm tin cơ sở mở rộng (EBBF)

* 1. Ví dụ minh họa

Ví dụ 1: Trong ví dụ cổ điển rằng FOD là X = {a, b, c}, hai BPA được cho là:

m1({a}) = 0.99, m1({c}) = 0.01, m1({∅}) = 0 (20)

m2({b}) = 0.99, m2({c}) = 0.01, m2({∅}) = 0. (21)

Trong ví dụ này, hàm khối lượng của tập rỗng là 0, chỉ ra rằng hàm khối lượng được phân bổ trong FOD toàn diện. Giá trị EBBF được tính như sau:

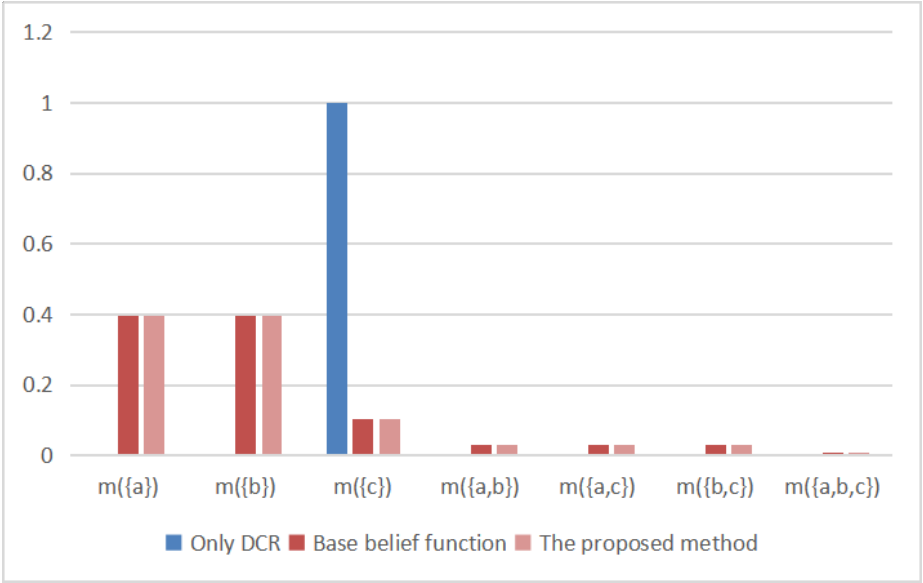
 (22)

Nó hoàn toàn giống với giá trị được tính với hàm niềm tin cơ bản. BPA của hai nhóm được sửa đổi tương ứng và kết quả được hiển thị trong Bảng 1.

**Bảng 1**. Chỉ định xác suất cơ bản đã sửa đổi (BPA) với phương pháp được đề xuất trong Ví dụ 1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| BPA | m({a}) | m({b}) | m({c}) | m({a,b}) | m({a,c}) | m({b,c}) | m({a,b,c}) | m({∅}) |
| m1(Ri) | 0.5664 | 0.0714 | 0.0764 | 0.0714 | 0.0714 | 0.0714 | 0.0714 | 0.0714 |
| m2(Ri) | 0.0714 | 0.5664 | 0.0764 | 0.0714 | 0.0714 | 0.0714 | 0.0714 | 0.0714 |

Cuối cùng, kết quả của việc hợp nhất dữ liệu được thể hiện trong Hình 2. Sau khi xem xét và nghiên cứu, phương pháp hợp nhất dữ liệu mà chúng tôi áp dụng ở đây là DCR thay vì quy tắc liên hợp TBM. Theo công thức của hai phương pháp, sự khác biệt giữa hai phương pháp có thể được phân tích như sau: Cả quy tắc liên hợp TBM và DCR chủ yếu xem xét ảnh hưởng của giao điểm, nhưng quy tắc liên hợp TBM không được chuẩn hóa, điều này làm cho tổng các giá trị niềm tin ít hơn hơn 1. Rõ ràng, kết quả như vậy là phản trực giác, vì vậy quy tắc liên từ TBM không phù hợp với các ví dụ trong bài viết này.



Hình 2. Kết quả hợp nhất của hai phương pháp trong Ví dụ 2

Kết quả hợp nhất cho thấy rằng quy tắc DCR không thể tạo ra kết quả logic trong ví dụ này, nhưng cả phương pháp trong bài báo này và phương pháp hàm niềm tin cơ sở đều có thể tạo ra kết quả trực quan. Cũng có thể thấy rằng với giả định FOD đầy đủ, phương pháp trong bài viết này có thể được rút gọn thành hàm niềm tin cơ sở.

Ví dụ 2: Giả sử rằng FOD là X = {a, b}, và hai BPA được cho là:

m1({a}) = 0.9, m1({∅}) = 0.1 (23)

m2({a}) = 0.9, m2({∅}) = 0.1 (24)

Tại thời điểm này, , cho thấy FOD không đầy đủ và ví dụ này nằm trong giả định FOD không đầy đủ. Quá trình tính toán được thể hiện trong H1

Bước 1: Từ FOD X = {a, b}, chúng ta có thể biết rằng tiềm năng của FOD là: N = 2

Bước 2: Đưa giá trị hàm khối lượng tập rỗng và N vào công thức và tính giá trị EBBF như sau:

neb(Ri) = = 0.3454 (25)

Bước 3: Sử dụng giá trị của hàm niềm tin để sửa đổi BPA của từng mẩu bằng chứng và tính toán như sau:

 (26)

 (27)

 (28)

 (29)

 (30)

 (31)

Bước 4: Đánh giá xem hàm khối lượng tập hợp rỗng có bằng 0 hay không và chọn một quy tắc kết hợp khác.

Trong ví dụ này, giá trị hàm khối lượng của tập hợp trống không phải là 0, vì vậy GCR được sử dụng để hợp nhất các BPA của hai tập hợp bằng chứng và thu được các kết quả sau:

m({a}) = 0.8053, m({b}) = 0.1195 (32)

m({a, b}) = 0.0398, m({∅}) = 0.0350 (33)

Kết quả cho thấy sự ủng hộ cao đối với mệnh đề a, phù hợp với tình hình thực tế của bằng chứng. Tuy nhiên, đồng thời, kết quả cũng cho b một mức độ hỗ trợ nhỏ nhất định, điều này cho thấy rằng chúng tôi chưa thu thập được tất cả các bằng chứng; vẫn có khả năng b đúng, tránh số liệu quá tuyệt đối.

Ví dụ 3: Trong FOD X = {a, b}, có hai bộ BPA xung đột cao, như sau:

m1({a}) = 0.9, m1({∅}) = 0.1 (34)

m2({b}) = 0.9, m2({∅}) = 0.1 (35)

Sử dụng phương pháp được đề xuất để tính giá trị hàm niềm tin, chúng ta có thể nhận được:

neb(Ri) = 0.3454 (36)

Sau đó, neb(Ri) được sử dụng để sửa đổi giá trị BPA của mỗi bằng chứng, như thể hiện trong Bảng 2.

**Bảng 2**. BPA được sửa đổi với phương pháp được đề xuất trong Ví dụ 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **BPA** | **m({a})** | **m({b})** | **m({a,b})** | **m({∅})** |
| m1(Ri) | 0.5229 | 0.1450 | 0.1450 | 0.1870 |
| m2(Ri) | 0.1450 | 0.5229 | 0.1450 | 0.1870 |

Các kết quả hợp nhất thu được bởi GCR được thể hiện trong Bảng 3. Dữ liệu cho thấy phương pháp được đề xuất đưa ra cùng một mức độ hỗ trợ cho a và b trong hai bộ dữ liệu hoàn toàn mâu thuẫn, đây là câu trả lời chính xác trong trực giác. Nó đã được chứng minh rằng phương pháp được đề xuất vẫn có thể cho kết quả tốt phù hợp với các sự kiện trong dữ liệu mâu thuẫn cao.

**Bảng 3**. Kết quả hợp nhất trong Ví dụ 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **BPA** | **m({a})** | **m({b})** | **m({a,b})** | **m({∅})** |
| Phương pháp đề xuất | 0.4548 | 0.4548 | 0.0554 | 0.0350 |

Ví dụ 4: Trong FOD X = {a, b, c}, các BPA như sau:

m1({a}) = 0.9, m1({a, b}) = 0.01, m1({∅}) = 0.09 (37)

m2({b}) = 0.01, m2({c}) = 0.91, m2({∅}) = 0.08. (38)

Sử dụng phương pháp được đề xuất để tính toán các giá trị của EBBF, chúng ta có thể nhận được:

m1 : neb(Ri)1 = 0.1448, m2 : neb(Ri)2 = 0.1446 (39)

Theo giá trị của hai EBBF, BPA của hai nhóm đã được sửa đổi cho phù hợp và kết quả được thể hiện trong Bảng 4

**Bảng 4.** BPA được sửa đổi bằng phương pháp được đề xuất trong Ví dụ 4

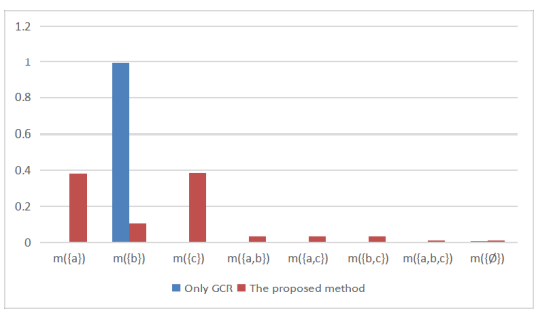
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **BPA** | **m({a})** | **m({b})** | **m({c})** | **m({a,b})** | **m({a,c})** | **m({b,c})** | **m({a,b,c})** | **m({∅})** |
| m1(Ri) | 0.4841 | 0. 0671 | 0. 0671 | 0. 0717 | 0. 0671 | 0. 0671 | 0.0671 | 0. 1088 |
| m2(Ri) | 0. 0670 | 0. 0717 | 0. 4890 | 0. 0670 | 0. 0670 | 0. 0670 | 0.0670 | 0. 1041 |

Sau đó, GCR được sử dụng cho phản ứng tổng hợp để thu được kết quả. Đồng thời, BPA từ bằng chứng trong bảng được hợp nhất trực tiếp với GCR và kết quả của hai phương pháp là được so sánh, như thể hiện trong Bảng 5 và Hình 3.

Trong trường hợp này, hai nhóm bằng chứng tương ứng ủng hộ cao cho a và c, nhưng kết quả hợp nhất chỉ sử dụng GCR sẽ gán mức hỗ trợ cao nhất cho đề xuất b, điều này rõ ràng là không hợp lý. Tuy nhiên, sau khi BPA được sửa đổi bằng phương pháp được đề xuất trong bài báo này, kết quả phản ứng tổng hợp

thu được bởi GCR đã hỗ trợ rất nhiều cho cả a và c. Hơn nữa, như là sự hỗ trợ cho c trong mảnh bằng chứng thứ hai lớn hơn một chút so với mảnh bằng chứng đầu tiên, kết quả hợp nhất cũng mang lại một chút lợi thế cho c. Điều này chỉ ra rằng trong trường hợp dữ liệu có tính xung đột cao, quyền câu trả lời trực quan không thể thu được đơn giản bằng cách sử dụng GCR, trong khi kết quả hợp lý và trực quan có thể thu được bằng cách sử dụng phương pháp trong bài viết này.

Các ví dụ trên xác minh rằng phương pháp được đề xuất tương thích với hàm niềm tin cơ sở trong FOD đầy đủ và xác minh tính khả thi và hiệu quả của phương pháp được đề xuất. Sau đây các ví dụ thảo luận về một số thuộc tính khác của phương pháp được đề xuất.



**Hình 3.** So sánh kết quả hợp nhất trong Ví dụ 4.

**Bảng 5.** Kết quả của hai phương pháp kết hợp trong Ví dụ 4.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Fusion Methods** | **m({a})** | **m({b})** | **m({c})** | **m({a,b})** | **m({a,c})** | **m({b,c})** | **m({a,b,c})** | **m({∅})** |
| Chỉ GCR | 0.0000 | 0.0660 | 0. 0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0. 0072 |
| Phương pháp đề xuất | 0.3831 | 0.1061 | 0.3851 | 0. 0354 | 0. 0338 | 0. 0338 | 0.0113 | 0.0113 |

**Ví dụ 5.** Trong FOD X = {a, b, c}, các BPA như sau:

m1({a}) = 0.7, m1({b}) = 0.1, m1({c}) = 0.1, m1({∅}) = 0.1 (40)

m2({a}) = 0.1, m2({b}) = 0.1, m2({c}) = 0.7, m2({∅}) = 0.1. (41)

Sử dụng phương pháp được đề xuất để tính giá trị của EBBF, chúng ta có thể nhận được:

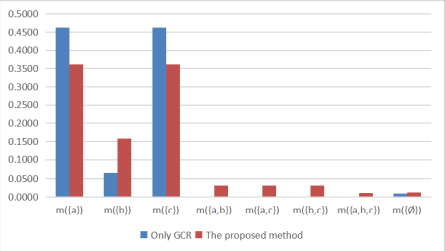
 (42)

Sau đó, BPA của từng mẩu bằng chứng đã được sửa đổi bằng cách sử dụng giá trị của hàm niềm tin được tính toán, và kết quả được thể hiện trong Bảng 6.

**Bảng 6.** BPA được sửa đổi bằng phương pháp được đề xuất trong Ví dụ 5

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **BPA** | **m({a})** | **m({b})** | **m({c})** | **m({a,b})** | **m({a,c})** | **m({b,c})** | **m({a,b,c})** | **m({∅})** |
| m1(Ri) | 0.3912 | 0. 1134 | 0. 1134 | 0. 0671 | 0. 0671 | 0. 0671 | 0.0671 | 0. 1134 |
| m2(Ri) | 0. 1134 | 0. 1134 | 0. 3912 | 0. 0671 | 0. 0671 | 0. 0671 | 0. 0671 | 0. 1134 |

Sau đó, GCR được sử dụng cho phản ứng tổng hợp để thu được kết quả. Đồng thời, BPA từ bằng chứng trong bảng được hợp nhất trực tiếp với GCR và kết quả của hai phương pháp là được so sánh, như thể hiện trong Bảng 7 và Hình 4



**Hình 4.** So sánh kết quả hợp nhất trong Ví dụ 5.

**Bảng 7.** Kết quả của hai phương pháp kết hợp trong Ví dụ 5.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Fusion Methods** | **m({a})** | **m({b})** | **m({c})** | **m({a,b})** | **m({a,c})** | **m({b,c})** | **m({a,b,c})** | **m({∅})** |
| Only GCR | 0.4620 | 0.0660 | 0. 4620 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0100 |
| The proposed method | 0.3620 | 0.1578 | 0.3620 | 0. 0316 | 0. 0316 | 0.0316 | 0.0105 | 0.0129 |

Theo Bảng 7, khi hàm khối lượng của từng tập hợp con phần tử đơn lẻ trong FOD không bằng 0, kết quả tổng hợp thu được bằng cách sử dụng phương pháp trong bài viết này để sửa đổi BPA và chỉ sử dụng GCR đều phản ánh rằng a và c có độ tin cậy cao như nhau. Chúng tôi đã tính thêm entropy các giá trị của hàm khối lượng bằng chứng ban đầu và hàm khối lượng được sửa đổi. Ở đây, chúng tôi đã thông qua mở rộng Deng entropy (EBEOW) [64], có thể tính toán giá trị entropy ở dạng không đầy đủ FOD. Theo tính toán, giá trị entropy Deng mở rộng của các BPA ban đầu là:

Eebeow(m)1 = 2.9237. (43)

Entropy Deng mở rộng của các BPA được sửa đổi bằng phương pháp đề xuất trong bài báo này là như sau:

Eebeow(m)2 = 3.8089. (44)

Rõ ràng, entropy tương ứng với kết quả tổng hợp của phương pháp được đề xuất tăng lên đáng kể, và sự gia tăng của entropy cho thấy rằng việc gán niềm tin bị phân tán hơn và sự không chắc chắn được tăng lên, đó là hậu quả của cuộc xung đột. Nói chung, khi đề xuất của một tập hợp con phần tử duy nhất khác không và dữ liệu rất mâu thuẫn, phương pháp được đề xuất trong phần này bài báo gán một phần niềm tin cho các phần tử đa tập hợp con khác, do đó giảm thiểu rủi ro.

**Ví dụ 6.** Giả sử FOD là X = {a, b, c} và hai BPA được cho là:

m1({a}) = 0.9, m1({a, b, c}) = 0.05, m1({∅}) = 0.05 (45)

m2({b}) = 0.9, m2({a, b, c}) = 0.05, m2({∅}) = 0.05 (46)

Theo phương pháp đề xuất, giá trị của hàm niềm tin tương ứng với bằng chứng m1 và m2 là được tính như sau:

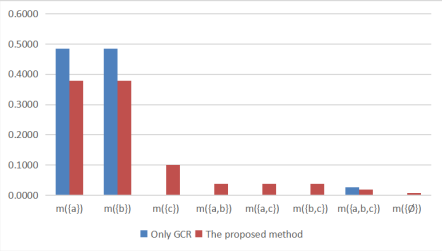
neb(Ri)1 = neb(Ri)2 = 0.1439 (47)

Sau đó, BPA của từng mẩu bằng chứng đã được sửa đổi bằng cách sử dụng giá trị của hàm niềm tin được tính toán, và kết quả được thể hiện trong Bảng 8.

**Bảng 8.** BPA được sửa đổi bằng phương pháp được đề xuất trong Ví dụ 6.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **BPA** | **m({a})** | **m({b})** | **m({c})** | **m({a,b})** | **m({a,c})** | **m({b,c})** | **m({a,b,c})** | **m({∅})** |
| m1(Ri) | 0.4852 | 0. 0669 | 0. 0669 | 0. 0669 | 0. 0669 | 0. 0669 | 0.0901 | 0. 0901 |
| m2(Ri) | 0.0669 | 0. 4852 | 0. 0669 | 0. 0669 | 0. 0669 | 0. 0669 | 0. 0901 | 0. 0901 |

Cuối cùng, kết quả hợp nhất thu được với GCR và kết quả hợp nhất với không sửa đổi BPA mà GCR thu được được trình bày trong Bảng 9 và Hình 5.



**Hình 5.** So sánh kết quả hợp nhất trong Ví dụ 6.

**Bảng 9.** Kết quả của hai phương pháp kết hợp trong Ví dụ 6.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Fusion Methods** | **m({a})** | **m({b})** | **m({c})** | **m({a,b})** | **m({a,c})** | **m({b,c})** | **m({a,b,c})** | **m({∅})** |
| Only GCR | 0.4853 | 0. 4853 | 0. 0000 | 0.0000 | 0.0000 | 0.0000 | 0.0270 | 0.0025 |
| Phương pháp đề xuất | 0.3796 | 0. 3796 | 0.0998 | 0.0380 | 0. 0380 | 0. 0380 | 0.0187 | 0.0081 |

Có thể thấy từ các kết quả trong Hình 5 rằng khi tất cả các hàm khối lượng của một tập hợp hoàn chỉnh là khác không, kết quả hợp nhất thu được bằng cách sử dụng phương pháp sửa đổi BPA được đề xuất trước và bằng cách chỉ sử dụng GCR, cả hai đều phản ánh rằng a và b có niềm tin cao như nhau, đây là một trực giác kết quả. Entropies niềm tin mở rộng của hàm khối lượng bằng chứng ban đầu và khối lượng được sửa đổi chức năng được tính toán thêm. Theo tính toán, entropy niềm tin mở rộng của BPA là Eebeow(m)1 = 2,8185 và entropy niềm tin mở rộng của BPA được sửa đổi theo phương pháp đề xuất là Eebeow(m)2 = 3,8488. Giá trị entropy tương ứng với kết quả hợp nhất của phương pháp này rõ ràng tăng. Do đó, khi hàm khối lượng của toàn bộ tập con phần tử của mỗi bằng chứng không bằng 0 và dữ liệu rất mâu thuẫn, phương pháp được đề xuất sẽ phân bổ một phần mức độ niềm tin cho những người khác đề xuất phụ, do đó làm giảm rủi ro.

**Ví dụ 7.** Trong FOD X = {a, b, c}, các BPA như sau:

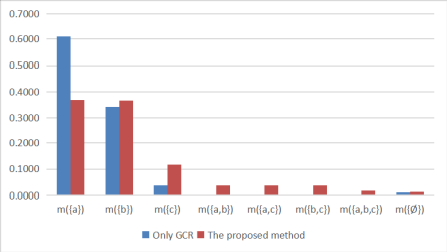
m1({a}) = 0.8, m1({b}) = 0.05, m1({c}) = 0.05, m1({∅}) = 0.1 (48)

m2({b}) = 0.8, m2({a, b, c}) = 0.1, m2({∅}) = 0.1. (49)

Sử dụng phương pháp được đề xuất để tính giá trị hàm niềm tin, chúng ta có thể nhận được:

neb(Ri)1 = neb(Ri)2 = 0.1450. (50)

Sau đó, GCR được sử dụng cho phản ứng tổng hợp để thu được kết quả. Đồng thời, BPA từ bằng chứng trong bảng được hợp nhất trực tiếp với GCR và kết quả của hai phương pháp là được so sánh, như thể hiện trong Bảng 10 và Hình 6.



**Hình 6.** So sánh kết quả hợp nhất trong Ví dụ 7.

**Bảng 10.** Kết quả của hai phương pháp kết hợp trong Ví dụ 7.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Fusion Methods** | **m({a})** | **m({b})** | **m({c})** | **m({a,b})** | **m({a,c})** | **m({b,c})** | **m({a,b,c})** | **m({∅})** |
| Only GCR | 0.6092 | 0.3427 | 0.0381 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0100 |
| Phương pháp đề xuất | 0.3699 | 0.3669 | 0.1176 | 0.0384 | 0.0384 | 0.0384 | 0.0176 | 0.0129 |

Khi hàm khối lượng tập hợp con đầy đủ của một tập hợp bằng chứng khác không, giá trị của tất cả các hàm khối lượng tập hợp con phần tử đơn của tập hợp bằng chứng khác là khác không và dữ liệu rất cao mâu thuẫn, kết quả hợp nhất chỉ sử dụng GCR chỉ mang lại mức độ tin cậy cao, trong khi kết quả nhiệt hạch sử dụng phương pháp trong bài báo này cho a và b độ hỗ trợ cao. Rõ ràng, kết quả

của phương pháp đề xuất phù hợp hơn với thực tế khách quan.

* 1. Ứng dụng cho dữ liệu nhân tạo

Để xác minh tính khả dụng và hiệu quả của phương pháp quản lý dữ liệu xung đột được đề xuất trong phần này, một ví dụ trong [65] được sử dụng để phân tích ví dụ và kết quả tính toán được so sánh với các phương pháp khác.

Xem xét vấn đề nhận dạng mục tiêu, người ta cho rằng có ba mục tiêu tiềm năng là được biểu diễn lần lượt là a, b và c. Theo phương pháp quản lý dữ liệu xung đột dựa trên EBBF đề xuất trong phần này, bước đầu tiên là lập mô hình bằng chứng thông tin không chắc chắn. Các báo cáo của năm cảm biến được mô hình hóa bởi BPA, và kết quả được thể hiện trong Bảng 11. Bảng 11 sửa đổi một chút

dữ liệu trong [65], gây ra sự mở rộng từ FOD đầy đủ sang FOD không đầy đủ. phân tích trực quan nhận thấy rằng báo cáo từ cảm biến thứ hai không nhất quán và xung đột với bốn cảm biến khác báo cáo. Ngoài ra, bốn cảm biến khác báo cáo rằng a có nhiều khả năng là mục tiêu tiềm năng nhất bởi vì giá trị hàm khối lượng của nó là lớn nhất và độ tin cậy là cao nhất.

**Table 11.** BPAs of the application example.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **BPA** | **m({a})** | **m({b})** | **m({c})** | **m({a,c})** | **m({∅})** |
| 1st sensor report: m1(·) | 0.41 | 0.29 | 0.2 | 0 | 0.1 |
| 2nd sensor report: m2(·) | 0 | 0.9 | 0.05 | 0 | 0.05 |
| 3rd sensor report: m3(·) | 0.58 | 0.07 | 0 | 0.15 | 0.2 |
| 4th sensor report: m4(·) | 0.55 | 0.1 | 0 | 0.15 | 0.2 |
| 5th sensor report: m5(·) | 0.6 | 0.1 | 0.0 | 0.2 | 0.1 |

Dựa trên dữ liệu trong Bảng 11, các giá trị EBBF của từng nhóm bằng chứng đã được tính toán như sau:

neb(Ri)1 = 0.1450 (51)

neb(Ri)2 = 0.1439 (52)

neb(Ri)3 = 0.1475 (53)

neb(Ri)4 = 0.1475 (54)

neb(Ri)5 = 0.1450 (55)

Theo các giá trị EBBF, các BPA đã được sửa đổi và dữ liệu đã sửa đổi của tất cả các BPA trong Bảng 11 được thể hiện trong Bảng 12.

**Bảng 12.** BPAs trong ứng dụng sau khi sửa đổi với phương pháp đề xuất.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| BPA | **m({a})** | **m({b})** | **m({c})** | **m({a,b})** | **m({a,c})** | **m({b,c})** | **m({a,b,c})** | **m({∅})** |
| m1(·) | 0.2569 | 0.2041 | 0.1597 | 0.0671 | 0.0671 | 0.0671 | 0.0671 | 0.1134 |
| m2(·) | 0.0669 | 0.4852 | 0.0901 | 0.0669 | 0.0669 | 0.0669 | 0.0669 | 0.0901 |
| m3(·) | 0.3337 | 0.0998 | 0.0677 | 0.0677 | 0.1365 | 0.0677 | 0.0677 | 0.1594 |
| m4(·) | 0.3199 | 0.1135 | 0.0677 | 0.0677 | 0.1365 | 0.0677 | 0.0677 | 0.1594 |
| m5(·) | 0.3449 | 0.1134 | 0.0671 | 0.0671 | 0.1597 | 0.0671 | 0.0671 | 0.1134 |

Hợp nhất dữ liệu bằng cách sử dụng quy tắc kết hợp tổng quát như sau:

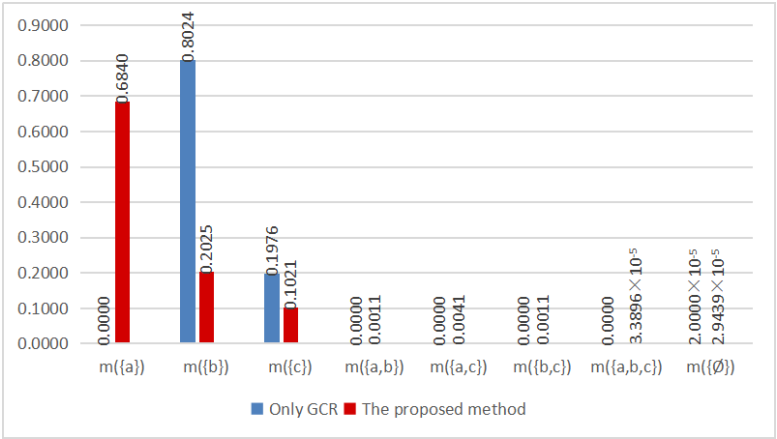
 (56)

Sau đó, GCR được sử dụng cho phản ứng tổng hợp để thu được kết quả. Đồng thời, các BPA từ bằng chứng trong bảng được hợp nhất trực tiếp với GCR và kết quả của hai phương pháp là được so sánh, như thể hiện trong Bảng 13 và Hình 7.

Kết quả tổng hợp cho thấy rằng sử dụng phương pháp được đề xuất trong phần này, có thể kết luận rằng a là mục tiêu đã xác định, phù hợp với kết quả phân tích trực quan. Tuy nhiên, kết quả của phản ứng tổng hợp GCR mà không có EBBF cho thấy b là mục tiêu đã xác định, trong khi khả năng a là mục tiêu là 0, đây rõ ràng không phải là kết quả trực quan. Qua so sánh trên, ta có thể thấy rõ rằng các kết quả hợp nhất dữ liệu xác minh tính hiệu quả của phương pháp được đề xuất trong phần này, có thể được sử dụng tốt trong quản lý dữ liệu xung đột trong kỹ thuật thực tế.

**Bảng 13.** Kết quả của hai phương pháp kết hợp trong ứng dụng dữ liệu nhân tạo.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Fusion Methods** | **m({a})** | **m({b})** | **m({c})** | **m({a,b})** | **m({a,c})** | **m({b,c})** | **m({a,b,c})** | **m({∅})** |
| Only GCR | 0.0000 | 0.8024 | 0.1976 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 2\*10-5 |
| The proposed method | 0.6840 | 0.2025 | 0.1021 | 0.0011 | 0.0041 | 0.0011 | 3.3896\*10-5 | 2.9439\*10-5 |



**Hình 7.** So sánh kết quả hợp nhất của ứng dụng.

* 1. Ứng dụng để phân loại

Để xác minh tính hiệu quả của phương pháp quản lý dữ liệu xung đột được đề xuất, một bộ dữ liệu thực để phân loại từ Kho lưu trữ máy học của Đại học California Irvine (UCI) đã được áp dụng trong ứng dụng này. Trong bộ dữ liệu mống mắt của UCI, có ba loại tròng đen: Setosa (a), Versicolor (b) và Virginia (c), mỗi loại chứa 50 mẫu. Mỗi mẫu chứa bốn thuộc tính, đó là chiều dài calax (SL), chiều rộng calax (SW), chiều dài cánh hoa (PL) và chiều rộng cánh hoa (PW). Một nhóm dữ liệu theo FOD không đầy đủ, được hiển thị trong Bảng 14, được thu thập sau quá trình tạo BPA trong [63].

Bảng 14. Kết quả mô hình hóa dữ liệu cho tròng đen có hàm khối lượng

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Attribute** | **m({a})** | **m({b})** | **m({c})** | **m({a,b})** | **m({a,c})** | **m({b,c})** | **m({a,b,c})** | **m({∅})** |
| SL | 0.6800 | 0.0000 | 0.0000 | 0.0610 | 0.0000 | 0.0000 | 0.1330 | 0.1260 |
| SW | 0.5030 | 0.0000 | 0.0000 | 0.0000 | 0.0100 | 0.0000 | 0.0000 | 0.4870 |
| PL | 0.9200 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0800 |
| PW | 0.8650 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.1350 |

Sử dụng phương pháp được đề xuất để tính toán các giá trị của EBBF, chúng ta có thể nhận được:

neb(Ri)SL = 0.1457, neb(Ri)SW = 0.1569, neb(Ri)PL = 0.1446, neb(Ri)PW = 0.1459 (57)

Sau đó, GCR được sử dụng để hợp nhất để thu được kết quả. Đồng thời, các BPA từ các bằng chứng trong bảng được hợp nhất trực tiếp với GCR và kết quả của hai phương pháp được so sánh, như thể hiện trong Bảng 15.

**Bảng 15**. Kết quả của hai phương pháp kết hợp trong ứng dụng để phân loại.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Fusion Methods** | **m({a})** | **m({b})** | **m({c})** | **m({a,b})** | **m({a,c})** | **m({b,c})** | **m({a,b,c})** | **m({∅})** |
| The proposed method | 0.9207 | 0.0363 | 0.0339 | 0.0031 | 0.0027 | 0.0027 | 0.0002 | 0.0005 |

Các kết quả hợp nhất được hiển thị trong Bảng 15. Kết quả cho thấy phương pháp này có hiệu suất hợp nhất tốt. Trong FOD không đầy đủ, lớp a chính xác vẫn được đưa ra một giá trị niềm tin rất cao, đó là kết quả trực quan. Thí nghiệm cho thấy phương pháp này cũng hợp lý và hiệu quả trong thế giới thực.

1. **Kết luận**

Trong FOD toàn diện, khi DCR được sử dụng để hợp nhất dữ liệu có tính mâu thuẫn cao, kết quả thường trái ngược với trực giác. Để giải quyết vấn đề này, một số học giả đưa ra chức năng niềm tin cơ bản để sửa đổi BPA nhằm loại bỏ tính tuyệt đối do xung đột mang lại. Theo giả thuyết FOD không đầy đủ, quy tắc kết hợp tổng quát có các vấn đề tương tự như DCR, nhưng chức năng niềm tin cơ bản không thể được áp dụng cho FOD không đầy đủ. Trong bài báo này, một hàm niềm tin cơ bản mở rộng dựa trên hàm niềm tin cơ bản và một phương pháp BPA sửa đổi tương ứng được đề xuất. Phương pháp này không chỉ tương thích với chức năng niềm tin cơ bản và có thể được sử dụng để quản lý dữ liệu xung đột trong FOD toàn diện, mà còn có thể được áp dụng hiệu quả trong FOD không đầy đủ. Sau khi sửa đổi BPA bằng EBBF, tính tuyệt đối do dữ liệu xung đột cao mang lại có thể được loại bỏ một cách hiệu quả. Phương pháp này không chỉ xem xét tiềm năng của FOD, mà còn xem xét các nguồn thông tin không chắc chắn không được tính đến bởi hàm niềm tin cơ sở và các phương pháp hiện có khác, bao gồm thông tin không chắc chắn do hàm khối lượng nonzero của tập rỗng và FOD không hoàn chỉnh mang lại. Để xác minh việc áp dụng phương pháp được đề xuất trong thực tế, bài báo này cũng thiết kế một phương pháp quản lý dữ liệu xung đột dựa trên EBBF và thảo luận và xác minh tính khả thi và hiệu quả của phương pháp thông qua một số ví dụ và ứng dụng.

Tuy nhiên, phương pháp được đề xuất trong bài báo này vẫn còn một số vấn đề mở đáng để thảo luận. Đầu tiên là vấn đề phức tạp tính toán. Phương pháp được đề xuất trong bài báo này có số lượng tính toán lớn như GCR, nhưng nó có thể được áp dụng cho nhiều tình huống hơn GCR. Nói tóm lại, phương pháp được đề xuất trong bài báo này không đơn giản, nhưng nó có hiệu quả. Sau đó, vì phương pháp này được đề xuất trong điều kiện thông tin không đầy đủ và trong FOD không đầy đủ, có nhiều nguồn thông tin không chắc chắn, hàm khối lượng tập rỗng tượng trưng được sử dụng trong phương pháp này. Trong các cải tiến tiếp theo, các nguồn thông tin không chắc chắn khác có thể được xem xét.

1. Phương pháp áp dụng lý thuyết chứng cứ vào việc hợp nhất dữ liệu mâu thuẫn
2. **Giới thiệu**

* Lý thuyết chứng cứ Dempster-Shafer (D-S) đã được áp dụng để xử lý thông tin không chắc chắn trong nhiều lĩnh vực như phân loại, gom cụm, chẩn đoán lỗi, hệ thống dựa trên kiến thức, chẩn đoán y tế, tổ hợp dữ liệu cảm biến, ra quyết định, phân tích rủi ro và nhiều lĩnh vực khác. Nhiều nỗ lực đã được đưa ra để giải quyết các vấn đề còn mở trong lý thuyết D-S. Trước hết, việc tạo ra phân bố xác suất cơ bản (BPA) là cơ sở để áp dụng lý thuyết D-S.
* Việc kết hợp các dữ liệu mâu thuẫn là một chủ đề nóng trong cả lý thuyết và thực tiễn. Lý thuyết D-S có thể tạo ra các kết quả đối nghịch với thực tế khi xử lý các dữ liệu mâu thuẫn một cách nghiêm trọng.
* Việc ra quyết định dựa trên hàm khối lượng cũng là một chủ đề nóng trong cả lý thuyết và thực tiễn. BPA được tạo ra không thể cung cấp trực tiếp xác suất của sự xuất hiện. Cách biến đổi BPA thành xác suất là một vấn đề lớn.
* Vấn đề về tính phức tạp tính toán cao. Tính phức tạp tính toán của quy tắc kết hợp Dempster là cao. Sau khi không gian được mở rộng, số sự kiện trong tập mũ tăng một cách mũ mà và độ phức tạp tính toán là cao.
* Hàm đánh giá tín hiệu, ví dụ như phương pháp độ mập độ tin cậy dựa trên entropy cho đo lường không chắc chắn. Bài báo này tập trung vào việc hợp nhất dữ liệu mâu thuẫn.
* Trong D-S theory, có nhiều phương pháp được áp dụng để quản lý thông tin không chắc chắn và dữ liệu mâu thuẫn . Lớp phương pháp đầu tiên là sửa đổi quy tắc kết hợp. Nhiều nhà nghiên cứu đã cố gắng sửa đổi quy tắc kết hợp trong D-S theory để xử lý dữ liệu mâu thuẫn cao.
* Trong [37] một quy tắc mới dựa trên khái niệm phân phối niềm tin chung được đề xuất. Những phương pháp này có thể mất tính toán học của quy tắc kết hợp Dempster và không đáp ứng được các luật kết hợp và luật hoán vị, điều này khiến chúng khó được áp dụng rộng rãi trong kỹ thuật. Lớp thứ hai là sửa đổi dữ liệu xung đột trước khi kết hợp dữ liệu. Trong [39] , entropy niềm tin được áp dụng để tiền xử lý dữ liệu xung đột trước khi kết hợp dữ liệu. Không chắc chắn của chứng cứ được mô hình hóa trước khi kết hợp chứng cứ trong [23].
* Lý luận chứng cứ (ER) cũng được phát triển trong lý thuyết D-S để mô hình và xử lý thông tin không chắc chắn [40-42]. Nó tăng cường quy tắc Dempster bằng cách xác định cách kết hợp nhiều bằng chứng hoàn toàn đáng tin cậy và giải quyết tốt các dữ liệu xung đột. Một thuật toán lý luận chứng cứ tổng quát đã được đề xuất trong [43], mở rộng thuật toán lý luận chứng cứ ban đầu. Trong [44], lý luận chứng cứ được áp dụng để thiết kế một phương pháp đánh giá quyết định đa thuộc tính mới, xem xét giá trị khoảng không đầy đủ. Mô hình lý luận chứng cứ cho cấu trúc niềm tin rời rạc được đề xuất trong [45]. Nó giải quyết vấn đề kết hợp các bằng chứng rời rạc không xung đột và khắc phục các kết quả mâu thuẫn khi kết hợp các bằng chứng mâu thuẫn nội hoặc ngoại tại.
* Một phương pháp đã được đề xuất để giải quyết vấn đề trọng số vượt quá giới hạn và sự phụ thuộc độ tin cậy của lý luận chứng cứ trong [46]. Nhiều khung nhận dạng dựa trên lý luận chứng cứ đã được đề xuất cho các ứng dụng thực tế. Trong [47], một phương pháp đánh giá quyết định đa thuộc tính mờ nhóm mới dựa trên tập mờ trực giác và lý luận chứng cứ được đề xuất.
* Trong [48], một phương pháp kết hợp có trọng số mới sử dụng lý luận chứng cứ cho nhiều bộ phân loại hoạt động với các đặc trưng khác nhau của mẫu đã được đề xuất. Thuật toán lý luận chứng cứ cũng được sử dụng trong nhiều lĩnh vực khác nhau, chẳng hạn như đánh giá an toàn trực tuyến [49] và hệ thống hỗ trợ ra quyết định [50]. Khung nhận dạng trong lý luận chứng cứ có hiệu suất tốt cho các hệ thống đòi hỏi độ chính xác cao. Để giải quyết các dữ liệu xung đột trong thông tin không chắc chắn, một số phương pháp cố gắng gán niềm tin ban đầu cho các sự kiện trong khung suy luận (FOD).
* Một chiến lược mới xem xét mức độ niềm tin ban đầu của một mệnh đề trong tập hợp mạnh của khung suy luận đã được đề xuất trong [51], nhưng nó có thể dẫn đến việc gán niềm tin phi tập trung. Việc gán niềm tin phi tập trung tăng mức độ niềm tin của các mệnh đề đa phần tử. Điều này gây ra sự tự tin được gán cho các tập không có ý nghĩa và không hữu ích cho việc kết hợp thông tin không chắc chắn và ra quyết định. Ngoài ra, các sự kiện trùng lắp có thể không xảy ra đồng thời từ quan điểm của lý thuyết xác suất. Do đó, việc gán niềm tin trên không gian tập hợp một cách trung bình có thể không hiệu quả. Để giải quyết vấn đề này, một phương pháp gán niềm tin xác suất cơ bản mới (bBPA) được đề xuất trong bài báo này.
* Phương pháp bBPA được đề xuất để xử lý thông tin không chắc chắn. Trong bBPA, niềm tin ban đầu được phân bổ trung bình cho các sự kiện cơ bản trong khung suy luận. Niềm tin ban đầu được phân bổ cho các sự kiện cơ bản có thể giới thiệu thông tin xác suất trước cho các phần tử. Trong một tình huống chưa biết, phân phối trung bình của niềm tin sẽ tối đa hóa entropy. Nguyên tắc tối đa entropy cho thấy khi entropy lớn nhất, sự mất mát có thể sẽ nhỏ nhất. Do đó, phân phối trung bình của niềm tin ban đầu giữa các phần tử mang lại thông tin trước hợp lý. Chiến lược này hoàn thành nhược điểm của BPA chỉ dựa trên một quan sát duy nhất. Bằng cách kết hợp thông tin xác suất trước đó, bBPA hoạt động tốt hơn trong việc hợp nhất dữ liệu xung đột và không xung đột.
* So với chiến lược trong [51], động lực và đóng góp của bBPA như sau.

Thứ nhất, trong các ứng dụng thực tế, mỗi sự kiện đối lập với nhau độc lập với nhau. Có thể chỉ có một phần tử một lúc. Các sự kiện đa tập hợp đại diện cho sự không chắc chắn trong việc xảy ra của các phần tử khác nhau.

Thứ hai, bBPA có độ phức tạp tính toán thấp hơn vì nó không gán giá trị tin cậy ban đầu trên toàn bộ không gian tập con của FOD, điều này có thể là một ưu điểm trong hệ thống thời gian thực.

Thứ ba, bBPA tránh việc gán giá trị tin cậy ban đầu trên toàn bộ không gian tập con, điều này sẽ giảm thiểu mất mát giá trị tin cậy trên nhiều tập con và có thể hữu ích cho việc đưa ra quyết định.

Thứ tư, bBPA có thể được giải thích tốt từ quan điểm của lý thuyết Bayes, và có thể tích hợp thông tin xác suất tiên nghiệm. Vì vậy, bBPA không chỉ phản ánh kết quả quan sát đơn lẻ, mà còn có thể tích hợp thông tin tiên nghiệm. Các ví dụ số và hai thí nghiệm trên UCI Machine Learning Repository được sử dụng để xác minh tính hiệu quả và hợp lý của phương pháp mới.

* Phần còn lại của bài báo được tổ chức như sau. Phần giới thiệu sơ lược được trình bày trong phần 2. Phần 3 đề xuất phương pháp gán phân bố xác suất cơ bản cho dữ liệu mâu thuẫn cùng với nhiều ví dụ số. Trong phần 4, bộ dữ liệu UCI Machine Learning Repository được sử dụng để xác minh tính hiệu quả của phương pháp mới trong bài toán phân loại. Kết luận được đưa ra trong phần 5.
  1. Quản lý xung đột với hàm tin cậy cơ bản.
* Quy tắc kết hợp Dempster có thể tạo ra các kết quả ngược đạo lý khi xử lý dữ liệu mâu thuẫn cao.
* Ví dụ 1 : (Zadeh 1986 [52]) Giả sử FOD là  và hai BPA được cho như sau:





Kết quả của việc hợp nhất dữ liệu bằng quy tắc kết hợp Dempster có thể dẫn đến kết quả ngược lại với logic khi xử lý dữ liệu mâu thuẫn cao. Cụ thể, kết quả của việc hợp nhất cho sự kiện b là m(b) = 1, điều này là không hợp lý. Theo logic, sự kiện b được hỗ trợ với ít niềm tin, nhưng niềm tin vào b lại được tăng lên với quy tắc kết hợp Dempster.

Để giải quyết vấn đề này, trong [51], hàm tin cậy cơ sở được đề xuất để sửa đổi BPA trước khi hợp nhất dữ liệu.

Gọi  là tập N giá trị có thể loại trừ lẫn nhau. Tập lũy thừa của  là , trong đó số phần tử là 

Nếu FOD hoàn thành, . Do đó, niềm tin cơ sở

hàm  được định nghĩa như sau [51]:

 (8)

trong đó là tập con  ngoại trừ tập rỗng . Sau đó

 được thông qua để sửa đổi BPA m ban đầu bằng cách tính toán

trung bình cộng [51]:

 (9)

Hàm tin cậy cơ bản gán cho mỗi tập con trong tập lực lượng một sự tin cậy ban đầu bằng nhau. Cần lưu ý rằng hàm tin cậy cơ bản gán tin cậy cho toàn bộ không gian tập con của FOD.

1. Bộ phận gán xác suất cơ bản mới để sửa đổi bằng chứng
   1. Bộ phận gán định lượng xác suất cơ bản mới

Trong một giả định thế giới đóng, đặt  là một tập hợp N giá trị có thể loại trừ lẫn nhau.



Bộ lũy thừa của Ω là , trong đó số phần tử là . Là một cách tiếp cận cải tiến của hàm niềm tin cơ sở trong [51], phép gán xác suất cơ bản cơ sở mới (bBPA), được ký hiệu là , được định nghĩa như sau:

 nếu  là một sự kiện tập con duy nhất (10)

0  nếu  không là một sự kiện tập con duy nhất

Phiên bản bBPA  mới được thiết kế để sửa đổi các BPA ban đầu nhằm giảm khả năng xung đột giữa các nhóm bằng chứng khác nhau. Hàm khối lượng đã sửa đổi có thể được tính toán và trình bày như sau:

 (11)

Thay vì chỉ định niềm tin cơ sở trong toàn bộ không gian tập hợp sức mạnh với (8), bBPA mới phân phối niềm tin cơ sở một cách trung bình trên yếu tố loại trừ lẫn nhau cơ bản, phù hợp với lý thuyết xác suất cổ điển.

Theo lý thuyết xác suất cổ điển, các sự kiện loại trừ lẫn nhau không thể xảy ra đồng thời. Do đó, bBPA phân phối trung bình độ tin cậy cho yếu tố loại trừ lẫn nhau cơ bản trong FOD. Vương và cộng sự. sử dụng 'những quả bóng màu trong một chiếc túi mờ đục' làm ví dụ để giải thích sơ đồ chuyển nhượng của niềm tin cơ sở trong [51]. Cụ thể, có ba quả bóng có màu sắc khác nhau và một túi màu đục, và ít nhất một quả bóng ở trong túi. Chúng tôi cho rằng, mỗi

lúc đó, một người có thể lấy ngẫu nhiên một quả bóng từ trong túi. Mỗi quả bóng có thể được chọn với xác suất bằng nhau tại một thời điểm. Giá trị khối lượng được phân phối trung bình giữa mỗi sự kiện cơ bản. Trong giả định thế giới đóng, mỗi sự kiện độc lập với nhau. Chỉ có một sự kiện xảy ra tại một thời điểm. Thông tin trước này có thể được lập mô hình với bBPA mới trong (11).

bBPA gán niềm tin ban đầu một cách trung bình cho sự kiện cơ bản là phần tử tập hợp con duy nhất. Ngoài tập rỗng, có N biến cố cơ bản trong một hệ kín. Do đó, bBPA được gán cho từng phần tử tập hợp con đơn lẻ là và bBPA cho mệnh đề tập hợp con không đơn lẻ là 0. Có thể giải thích rằng xác suất ưu tiên ban đầu được gán cho từng phần tử cơ bản để khắc phục nhược điểm của kết hợp Dempster quy tắc trong việc hợp nhất dữ liệu xung đột.

* 1. **Khả năng tương thích với lý thuyết Bayesian**

Trong lý thuyết Bayes, xác suất sau tỷ lệ thuận với xác suất trước. Dữ liệu mới có thể được lấy từ thông tin có giá trị để sửa đổi xác suất trước đó. bBPA gán trung bình niềm tin ban đầu cho các sự kiện cơ bản, có nghĩa là sự kết hợp giữa thông tin xác suất trước đó và thông tin quan sát. Trong một tình huống không xác định, phân phối trung bình của niềm tin sẽ tối đa hóa entropy.

Nguyên lý entropy cực đại cho thấy khi entropy lớn nhất thì tổn thất có thể xảy ra sẽ nhỏ nhất. Do đó, bBPA thể hiện nguyên tắc entropy cực đại. Hơn nữa, nếu chúng ta có hiểu biết nhất định về sự kiện chưa biết, tương ứng với thông tin xác suất trước của sự kiện, chúng ta có thể sử dụng xác suất trước để cập nhật phân phối xác suất ban đầu.

Cụ thể, đối với mỗi sự kiện cơ bản, (10) có thể được cập nhật như sau:

 , nếu  là một sự kiện tập con duy nhất (12)

0 , nếu  không là một sự kiện tập con duy nhất

trong đó đại diện cho xác suất tiên nghiệm của. Lưu ý rằng tổng xác suất trước của các sự kiện cơ bản phải bằng 1, được biểu thị như sau:

 (13)

Theo (12), chúng tôi kết hợp BPA với thông tin trước, điều này sẽ hữu ích cho việc tổng hợp thông tin và ra quyết định trong điều kiện không chắc chắn. Nói chung, chúng tôi xem xét toàn bộ hệ thống. Lúc đầu, chúng tôi không có thông tin về xác suất xảy ra của sự kiện cơ bản trong hệ thống. Xem xét nguyên tắc entropy tối đa, chúng tôi sử dụng bBPA để sửa đổi BPA. Khi chúng ta có đủ thông tin tiên nghiệm về sự kiện, (12) có thể được sử dụng để giới thiệu thông tin tiên nghiệm cho BPA. Trong trường hợp thông tin trước có ảnh hưởng mạnh đến kết quả, chúng ta có thể đặt ngưỡng xác suất trước dựa trên bài toán cụ thể. Một ví dụ về quy trình này sẽ được giới thiệu trong Phần 4. Các ví dụ sau đây sẽ giải thích và phân tích những ưu điểm của việc sử dụng bBPA theo nguyên tắc entropy tối đa trong việc xử lý dữ liệu có độ xung đột cao, trong đó bBPA có thể được coi là thông tin trước dưới dạng Xác suất Bayes.

* 1. **Ví dụ số**
     1. **Hợp nhất dữ liệu xung đột**

Các ví dụ bằng số sau đây được đề xuất để xác minh tính hiệu quả của phương pháp bBPA trong việc xử lý dữ liệu có tính xung đột cao. Kết quả được so sánh với phương pháp trong [51]. Trước hết, hãy nhớ lại trường hợp hợp nhất dữ liệu xung đột kinh điển do Zadeh [52] đề xuất.

Ví dụ 2 Xác định FOD là  hai BPA được cho như sau:





Với hàm niềm tin cơ sở mb trong (8), sau đây

chức năng niềm tin cơ sở được xây dựng:



=

Do đó, kết quả hợp nhất của dữ liệu xung đột sau khi sửa đổi dựa trên chức năng niềm tin cơ sở như sau:

,



Dựa trên phương pháp mới trong (10), hàm bBPA được đề xuất cho trường hợp này sẽ được đưa ra như sau:



Theo (11), kết quả của bằng chứng dựa trên bBPA

sửa đổi là:









Do đó, với quy tắc kết hợp Dempster, dữ liệu xung đột đã sửa đổi có thể được hợp nhất như sau:



Trong cả hai BPA ban đầu, niềm tin về giống như niềm tin của .Tuy nhiên, vì bằng chứng đầu tiên gán niềm tin vào nên mệnh đề có thể có niềm tin không chắc chắn cao hơn mệnh đề .Từ quan điểm này, kết quả hợp nhất dựa trên bBPA mới hợp lý hơn trong trường hợp này vì nó bắt được nhiều độ không đảm bảo hơn so với phương pháp ban đầu. Ngoài ra, không có niềm tin vào

các sự kiện không chắc chắn như và ,điều này rất hữu ích cho việc gán niềm tin tập trung. Phương pháp mới có thể giải quyết thông tin không chắc chắn chính xác hơn bằng cách nắm bắt nhiều thông tin không chắc chắn hơn trong BPA. Trên hết, phương pháp dựa trên bBPA mới góp phần hội tụ niềm tin cho mệnh đề tập hợp con duy nhất, giúp ích cho việc ra quyết định trong các ứng dụng thực tế.

Ví dụ 3 Giả sử rằng FOD là Ω = {a, b} và hai BPA được cho là:

,



Với hàm niềm tin cơ sở trong (8), các hàm niềm tin cơ sở sau được xây dựng:



BPA đã sửa đổi có thể được hợp nhất bằng cách sử dụng quy tắc kết hợp Dempster và kết quả như sau:



Với phương pháp được đề xuất trong (10), bBPA mới cho BPA

sửa đổi nên là:



Với (11), kết quả của phép biến đổi dựa trên bBPA là:









Với quy tắc kết hợp Dempster, xung đột được sửa đổi

dữ liệu có thể được hợp nhất như sau:

0.5

Kết quả cho thấy rằng mỗi và  chiếm 50% niềm tin, điều này là hợp lý. Đồng thời, so với phương pháp cũ, bBPA không gán niềm tin cho .

Ví dụ 4 Giả sử rằng FOD là  và hai BPA được cho là:



Kết quả của phương pháp đề xuất so với phương pháp cũ được thể hiện trong Bảng 1. Niềm tin được gán nhiều hơn cho các mệnh đề {a} và {c}, vượt trội so với phương pháp cũ và hữu ích cho việc ra quyết định.

Ví dụ 5 Một ví dụ đặc biệt khi không có phần tử tập hợp con nào.



Với (10), bBPA mới để sửa đổi BPA phải là:



Sử dụng (11), kết quả của sửa đổi dựa trên bBPA là:









Với quy tắc kết hợp Dempster, dữ liệu xung đột đã sửa đổi có thể được hợp nhất như sau:





Dựa trên phương pháp mới, một mức độ niềm tin cao đã được trao cho phần tử {c} xuất hiện trong mỗi đề xuất của hai nhóm BPA. Đồng thời, nó đã cung cấp một trọng lượng nhất định cho cả {a} và {b} để làm cho kết quả có kết quả hội tụ trên phần tử tiêu cự đơn.

* + 1. **Phản ứng tổng hợp dữ liệu mà không có xung đột**

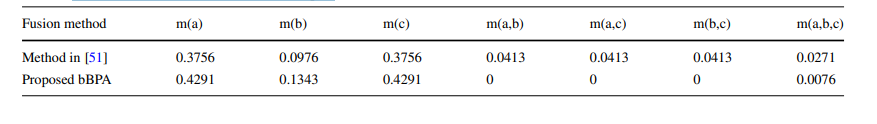
Để thể hiện tính ưu việt của phương pháp đề xuất so với phương pháp ban đầu, một số ví dụ về sự hợp nhất dữ liệu mà không có xung đột giữa các phần bằng chứng khác nhau được thiết kế như sau.

Ví dụ 6 Giả sử rằng FOD là Ω = {a, b}, hai BPA được cho như sau:





**Bảng 1** Kết quả của hai quy tắc kết hợp của Ví dụ 5



Theo hàm niềm tin cơ sở trong (8),



Sau đó, việc sửa đổi hai BPA dựa trên (9) được hiển thị như sau:



trong đó i = 1, 2. BPA đã sửa đổi có thể được hợp nhất theo quy tắc kết hợp Dempster và kết quả như sau:



Theo kết quả hợp nhất, niềm tin vào {a} và {b} lần lượt là 85,71% và 10,71%. Cần lưu ý rằng niềm tin 3,57% được gán cho mệnh đề {a, b}, điều này làm suy yếu niềm tin vào một tập hợp con duy nhất và không hữu ích cho việc ra quyết định. Mặc dù niềm tin vào {a, b} là nhỏ trong trường hợp này, tuy nhiên, mỗi mẩu bằng chứng nhỏ có thể được coi là thông tin có giá trị đều quan trọng trong quyết định

làm.

Áp dụng bBPA mới để giải quyết trường hợp hợp nhất dữ liệu tương tự.

Với (10), bBPA mới để sửa đổi BPA phải là:



Sau đó, dựa vào (11), kết quả của phép biến đổi dựa trên bBPA là:



trong đó i = 1, 2. Cuối cùng, với quy tắc kết hợp Dempster, kết quả hợp nhất dựa trên bBPA là:



Như có thể thấy từ các kết quả, phương pháp dựa trên bBPA có mức độ tin cậy cao hơn trên {a} so với phương pháp có (8)-(9). Phương pháp dựa trên bBPA không làm mất niềm tin vào mệnh đề {a, b}, có nghĩa là ít mất thông tin hơn. Kết quả cho thấy rằng ngay cả khi hợp nhất dữ liệu mà không có xung đột giữa các BPA khác nhau, phương pháp được đề xuất có thể mang lại kết quả hợp nhất tốt hơn so với phương pháp cũ.

Ví dụ 7 Giả sử FOD là , một số BPA được cho như sau:



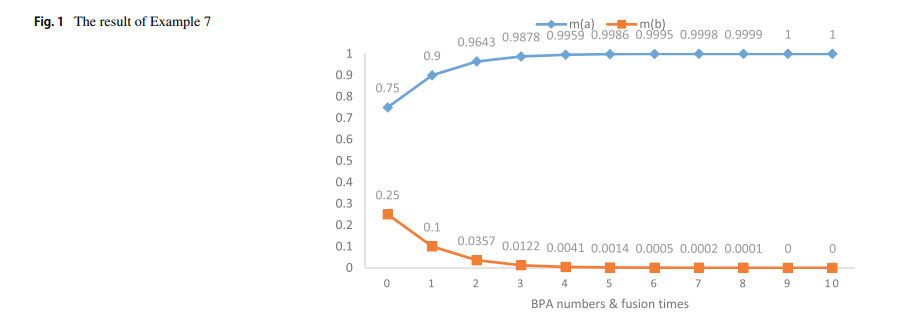
trong đó i đại diện cho số lượng BPA tuần tự để hợp nhất dữ liệu (i=1, 2, 3...). Đây là trường hợp đặc biệt, các chuyên gia đều đưa ra đánh giá như vậy. Áp dụng bBPA mới để giải quyết trường hợp hợp nhất dữ liệu này. Với (10), bBPA mới để sửa đổi BPA phải là:

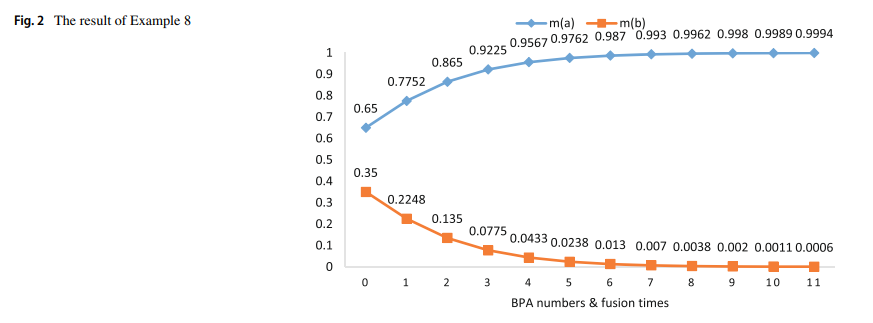


Sau đó, dựa vào (11), kết quả của phép biến đổi dựa trên bBPA là:



Cuối cùng, với quy tắc kết hợp Dempster, kết quả hợp nhất dựa trên bBPA có thể được tính toán. Các kết quả được hiển thị trong Hình 1. Có thể thấy từ Hình 1, khi ngày càng có nhiều bằng chứng được hợp nhất, m(a) và m(b) nhanh chóng phát triển thành hai thái cực. Khi bốn phần bằng chứng về dữ liệu được hợp nhất, giá trị của m(a) gần bằng 1 và giá trị của m(b) gần bằng 0. bBPA cho thấy hiệu suất tốt trong việc xử lý dữ liệu không xung đột.



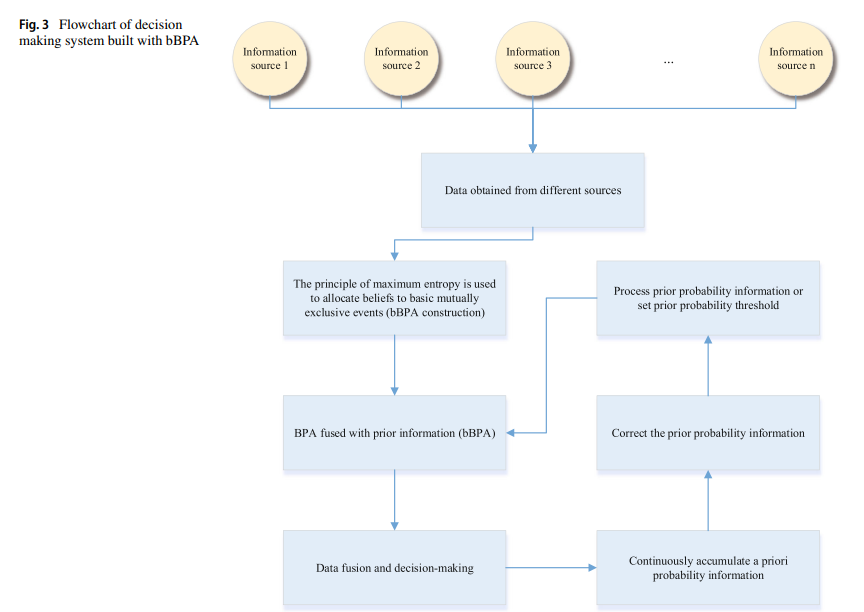


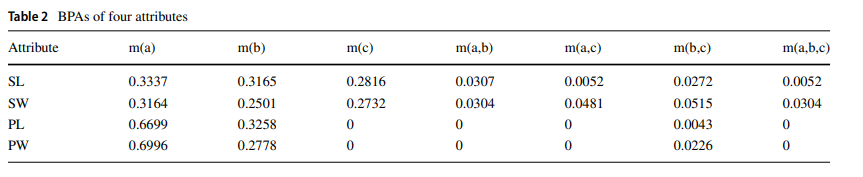
Ví dụ 8 Giả sử FOD là , một số BPA được cho như sau:



trong đó i đại diện cho số lượng BPA tuần tự để hợp nhất dữ liệu (i=1, 2, 3...).

Trong trường hợp này, các chuyên gia cung cấp cùng một loại BPA. Với phương pháp của chúng tôi, kết quả thử nghiệm được hiển thị trong Hình 2. Giá trị của m(a) nhanh chóng hội tụ về 1 và giá trị của m(b) nhanh chóng hội tụ về 0. bBPA có thể tích hợp các ý kiến từ các chuyên gia khác nhau, điều này khiến độ tin cậy của a cao hơn nhiều so với b. Kết quả là hợp lý khi độ tin cậy của mỗi chuyên gia là như nhau.





* + 1. **bBPA với thông tin trước**

Trong phần này, một số ví dụ được sử dụng để giới thiệu phương pháp theo nguyên lý Bayes. Cơ sở của bBPA phù hợp với xác suất Bayes.

Ví dụ 9 Giả sử FOD là , hai BPA được cho như sau:



Trong khi đó, chúng tôi giả định rằng độ tin cậy được gán cho a là 70% và độ tin cậy của b là 30%, được ký hiệu là:



Với quy tắc kết hợp Dempster và không có bBPA, ta sẽ được kết quả kết hợp:



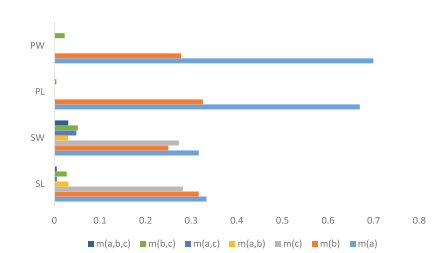
Dựa trên (12), bBPA mới đến từ xác suất biến đổi BPA trước đó phải là:



Do đó, dựa trên (11), kết quả của sửa đổi dựa trên bBPA là:

,





Hình 4 Phân phối BPA trong phân loại Iris

Cuối cùng, với quy tắc kết hợp Dempster, dữ liệu xung đột đã sửa đổi có thể được hợp nhất như sau:



Với sự hỗ trợ của các thông tin trước đó, có một mức độ tin cậy cao hơn cho mệnh đề a. Điều này có thể được giải thích là bản ghi lịch sử có tác động tích cực đến việc lựa chọn a.

Ví dụ 10 Xác định rằng FOD là , hai BPA được cho như sau:



Nhớ lại ví dụ cổ điển này một lần nữa, ở đây, chúng tôi có thông tin được cung cấp trước. Xác suất trước của sự kiện được đưa ra như sau:



Sử dụng (12) và (11), kết quả của sửa đổi dựa trên bBPA là:





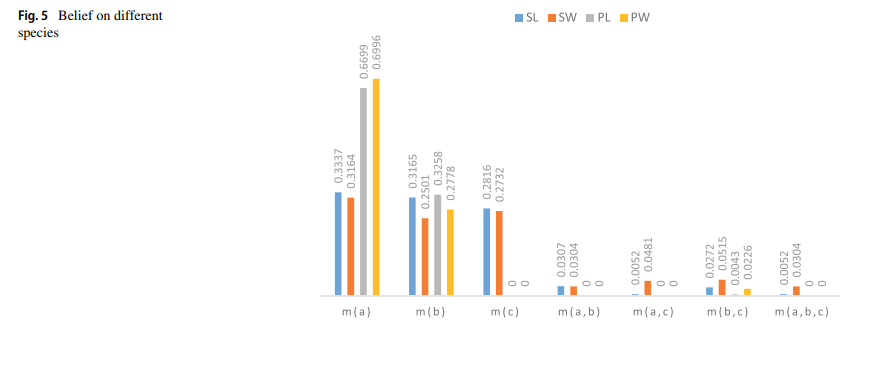
Do đó, với quy tắc kết hợp Dempster, dữ liệu xung đột đã sửa đổi có thể được hợp nhất như sau:



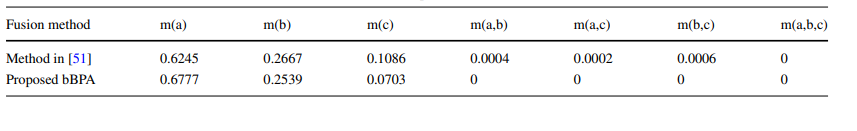
Có thể thấy rằng trên cơ sở xác suất tiên nghiệm, phương pháp đề xuất cũng có thể xử lý các dữ liệu mâu thuẫn một cách hiệu quả. Thông tin trước đó có thể được tích hợp vào phương pháp được thiết kế.

1. **Ứng dụng trong bài toán phân lớp**

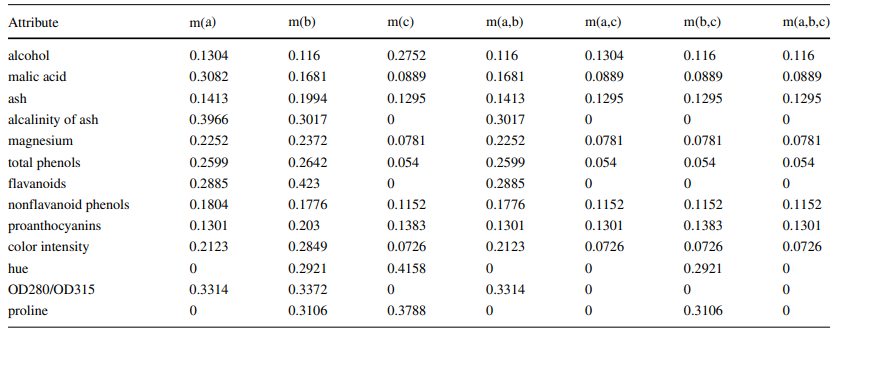
Trong phần này, chúng tôi sử dụng tập dữ liệu từ UCI Machine Learning Repository để xác minh tính hiệu quả của phương pháp bBPA mới so với phương pháp ban đầu trong [51].



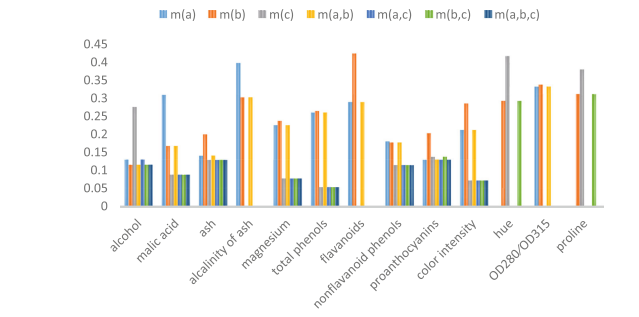
Bảng 3 Kết quả của các phương pháp hợp nhất dữ liệu khác nhau cho thí nghiệm phân loại Iris



Bảng 4 BPA của 13 thuộc tính



Hình 6 Phân bố BPA trong thí nghiệm 2



* 1. **Phương pháp hợp nhất dữ liệu dựa trên bBPA mới**

Lưu đồ của phương pháp hợp nhất dữ liệu dựa trên phương pháp bBPA mới được hiển thị trong Hình 3.

**Bước 1** Nguồn thông tin tạo ra một loạt các BPA có thể chứa dữ liệu xung đột.

**Bước 2** bBPA mới được xây dựng theo số lượng các sự kiện cơ bản trong đó niềm tin cơ sở được gán cho các sự kiện loại trừ lẫn nhau.

**Bước 3** Sửa đổi BPA ban đầu bằng bBPA để xử lý trước dữ liệu xung đột.

**Bước 4** Hợp nhất dữ liệu bằng quy tắc kết hợp Dempster.

**Bước 5** Ra quyết định dựa trên kết quả hợp nhất dữ liệu trong phân loại.

**Bước 6** Cập nhật thông tin xác suất trước theo thông tin xác suất tích lũy.

**Bước 7** Đặt ngưỡng cho xác suất trước để nó không quá lớn.

**Bước 8** Sử dụng thông tin trước để xử lý các BPA sau.

* 1. Thí nghiệm 1: Phân loại tập dữ liệu Iris

Có ba loài (Setosa(a), V ersicolor(b), V irginica(c)) trong bộ dữ liệu Iris với bốn thuộc tính (SL, SW, PL và PW) và mỗi loài chứa 50 cá thể.

Các BPA được tạo ra trong [51] được thông qua, trong đó Wang et al. chọn ngẫu nhiên 40 trường hợp từ mỗi loài và tạo các số mờ tam giác [53] của bốn thuộc tính. 10 trường hợp còn lại là các bộ kiểm tra. Vương và cộng sự. chọn ngẫu nhiên một cá thể từ loài Setosa (a) làm bộ thử nghiệm và tạo BPA. Kết quả được thể hiện trong Bảng 2. Sự phân bố BPA được thể hiện trong Hình. 4 và 5.

Hình 4 cho thấy rõ ràng sự phân bố BPA được tạo ra bởi các thuộc tính khác nhau. Màu sắc khác nhau đại diện cho các sự kiện khác nhau. Trong Hình 5, sự phân bố BPA do các thuộc tính khác nhau tạo ra được thể hiện từ một góc nhìn khác. Màu sắc khác nhau đại diện cho các thuộc tính khác nhau.

Trong Hình 4, các thuộc tính SL và SW của hoa được đưa ra đánh giá tương tự đối với loài {a}, {b} và {c}, nhưng các thuộc tính PL và PW cho loài {a} và loài {b } với sự hỗ trợ niềm tin cao hơn. Trong Hình 5, có thể thấy rõ rằng bốn thuộc tính khác nhau đều có mức độ tin cậy cao nhất đối với loài {a}. Vì vậy, thật hợp lý khi tin rằng loài Setosa có nhiều khả năng là loại hoa được công nhận. Cuối cùng, so sánh kết quả sử dụng phương pháp trong [51] và phương pháp bBPA được đề xuất, kết quả với các phương pháp tiền xử lý dữ liệu khác nhau được trình bày trong Bảng 3.

Như có thể thấy từ Bảng 3, tất cả các phương pháp đều có thể nhận ra cá thể thử nghiệm có khả năng là loài Setosa ({a}) với độ tin cậy trên 60%, phù hợp với tình hình thực tế. Phương pháp bBPA mới mang lại mức độ tin cậy cao hơn đối với loài {a}. Ngoài ra, phương pháp đề xuất không làm mất niềm tin vào đề xuất có nhiều loài, điều này cho thấy tính ưu việt của bBPA trong việc giải quyết vấn đề phân loại.

* 1. **Thí nghiệm 2: Phân loại tập dữ liệu rượu vang**

Thử nghiệm phân loại của bộ dữ liệu rượu vang được thông qua để

xác minh phương pháp bBPA mới. Bộ dữ liệu rượu bao gồm 3 loại rượu khác nhau (ký hiệu là {a}, {b} và {c}) với 13 thuộc tính .Các BPA được tạo thông qua [51] được hiển thị trong Bảng 4. Hình 6 cho thấy sự phân bố BPA được tạo bởi 13 thuộc tính khác nhau

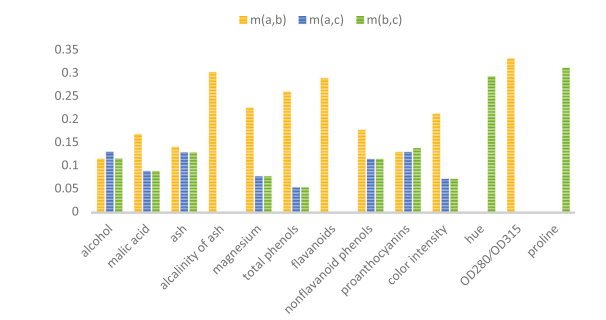
Theo Hình 6, không thể hình dung rõ ràng tất cả các mối quan hệ nội tại giữa các giống nhưng có thể nhận thấy mức độ tin tưởng đối với giống {a} và {b} cao hơn giống {c}. Bằng trực giác, có thể giả định rằng các giống {a} và {b} có nhiều khả năng là giống tiềm năng được công nhận hơn so với giống {c}.

Trong Hình 7, chỉ tập hợp con {a, c}, {b, c} và {a, b} được giữ lại. Có thể thấy rõ rằng trong hầu hết các thuộc tính, tập con {a, b} có mức độ tin cậy cao hơn. Trong Hình 8, một phần khác của thuộc tính bị ẩn, chỉ tập hợp con {a}, {b} và {c} được giữ lại. Có thể thấy rằng hầu hết các thuộc tính cung cấp mức độ tin cậy cao hơn trên {a} và {b}. Do đó, so với {c}, các loại {a} và {b} sẽ có nhiều khả năng trở thành giống tiềm năng được công nhận hơn.

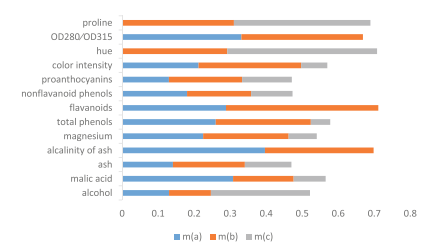
Hình 9 cho thấy sự phân bổ BPA của các giống {a} và {b}, trong đó các BPA của {c}, {a, b} và {a, b, c} bị ẩn vì chúng không phải là vấn đề chính đối với giống tiềm năng đã được công nhận. Có thể thấy rõ, trong 13 thuộc tính thì giống {b} có mức độ tin tưởng cao hơn giống {a}, còn thuộc tính {b, c} có mức độ tin cậy cao hơn {a, c}. Do đó, {b} có nhiều khả năng là biến thể được công nhận trong vấn đề.

Cuối cùng, kết quả của sự hợp nhất dữ liệu được thể hiện trong Bảng 5. Cả ba phương pháp đều chỉ định mức độ tin cậy cao hơn cho biến thể {b}. 9 trong số 13 BPA cho rằng {b} có niềm tin cao hơn {a}. Kết quả tổng hợp dữ liệu cho thấy {c} không phải là giống được công nhận. Một số niềm tin được gán cho {a}, phù hợp với ứng dụng thực tế. Do đó, phương pháp bBPA mới là hợp lý và hiệu quả.

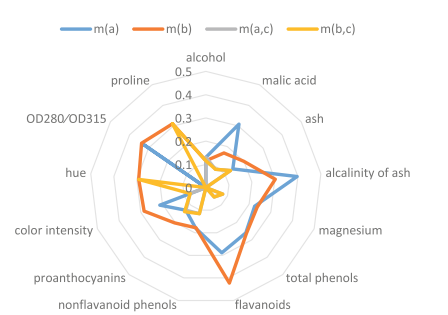
Hình 7. Phân bố của tập con có hai phần tử trong thí nghiệm 2



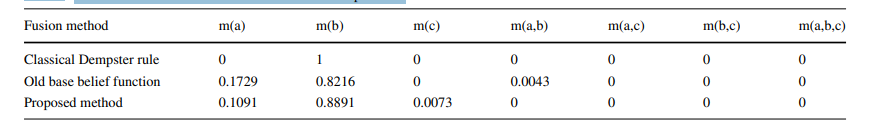
Hình 8 Phân phối một tập hợp con BPA trong thí nghiệm 2



Hình 9 Phân phối hỗ trợ cho các sự kiện bằng các bằng chứng khác nhau



Bảng 5 Kết quả của các quy tắc kết hợp khác nhau của thí nghiệm Wine



1. **Kết Luận**

Quy tắc kết hợp Dempster truyền thống có thể tạo ra kết quả phản trực giác trong khi xử lý dữ liệu có tính xung đột cao. Hàm niềm tin cơ sở có thể là một phương pháp để giải quyết vấn đề này. Trong bài báo này, một phương pháp hàm niềm tin cơ sở mới có tên là bBPA được đề xuất. Những điểm nổi bật của bBPA như sau. Trước hết, nó có độ phức tạp tính toán thấp hơn vì bBPA không cần gán niềm tin cơ sở trên toàn bộ không gian tập hợp sức mạnh, đây có thể là một ưu điểm trong hệ thống thời gian thực. Thứ hai, bBPA phân phối niềm tin trung bình cho các sự kiện cơ bản, điều này phù hợp với lý thuyết xác suất cổ điển. Cuối cùng, bBPA có thể thể hiện và kết hợp thông tin trước đó và tránh làm tăng niềm tin vào nhiều tập hợp con không hữu ích cho việc ra quyết định.

bBPA dựa trên sự kiện cơ bản. Phần còn lại của các sự kiện không chắc chắn trong tập hợp sức mạnh của FOD có thể không hữu ích cho việc ra quyết định. Vì vậy, chúng tôi chỉ gán niềm tin cơ sở ban đầu cho sự kiện cơ bản. Thông tin trước đó có thể được lập mô hình trong hợp nhất dữ liệu với bBPA. Kinh nghiệm hiện có từ các chuyên gia có thể được áp dụng để đối phó với sự kết hợp thông tin mâu thuẫn. Ngoài ra, bBPA còn làm giảm mức độ tin cậy của mệnh đề có nhiều yếu tố, giúp phương pháp dựa trên bBPA hiệu quả hơn trong các bài toán ra quyết định. Các ví dụ và thử nghiệm cho thấy rằng phương pháp được đề xuất có thể xử lý hiệu quả việc hợp nhất dữ liệu xung đột.

Các công việc tiếp theo có thể tập trung vào việc áp dụng phương pháp đề xuất trong việc ra quyết định cho các vấn đề thực tế. Quan trọng hơn, hiện tại, làm thế nào để gán niềm tin của một mệnh đề có nhiều yếu tố cho một tập hợp con duy nhất vẫn là một vấn đề mở. Công việc sau đây của phương pháp ra quyết định dựa trên bBPA nên xem xét vấn đề này.

1. **Chương 2: Các luật hợp nhất tri thức**
2. Dempster’s combination rule

Ví dụ 1: Cho Ω = {A,B,C}

m1(A) = 0.7, m1(B) = 0.2, m1(C) = 0.1

m2(A) = 0.3, m2(B) = 0.2, m2(C) = 0.5

Với Ω = {A,B,C}

|  |  |  |
| --- | --- | --- |
|  | **m1** | **m2** |
| **ABC** | 0.0000 | 0.0000 |
| **AB** | 0.0000 | 0.0000 |
| **AC** | 0.0000 | 0.0000 |
| **A** | 0.7000 | 0.3000 |
| **BC** | 0.0000 | 0.0000 |
| **B** | 0.2000 | 0.2000 |
| **C** | 0.1000 | 0.5000 |
| **Æ** | 0.0000 | 0.0000 |

Hệ số chuẩn hóa k:

|  |  |
| --- | --- |
| **BC=** |  |
| m1(ABC)\*m2() | 0 |
| m1(AB)\*m2(C) | 0 |
| m1(AB)\*m2() | 0 |
| m1(AC)\*m2(B) | 0 |
| m1(AC)\*m2() | 0 |
| m1(BC)\*m2(A) | 0 |
| m1(BC)\*m2() | 0 |
| m1(A)\*m2(B) | 0.14 |
| m1(A)\*m2(C) | 0.35 |
| m1(A)\*m2(BC) | 0 |
| m1(A)\*m2() | 0 |
| m1(B)\*m2(A) | 0.06 |
| m1(B)\*m2(C) | 0.1 |
| m1(B)\*m2(AC) | 0 |
| m1(B)\*m2() | 0 |
| m1(C)\*m2(A) | 0.03 |
| m1(C)\*m2(B) | 0.02 |
| m1(C)\*m2(AB) | 0 |
| m1(C)\*m2() | 0 |
| m1()\*m2(ABC) | 0 |
| m1()\*m2(AB) | 0 |
| m1()\*m2(AC) | 0 |
| m1()\*m2(BC) | 0 |
| m1()\*m2(A) | 0 |
| m1()\*m2(B) | 0 |
| m1()\*m2(C) | 0 |
| m1()\*m2() | 0 |
| **K** | **0.7** |

Hợp nhất 2 hàm khối lượng

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ABC** | **AB** | **AC** | **BC** | **A** | **B** | **C** |
| m1(ABC)\*m2(ABC)  0 | m1(ABC)\*m2(AB)  0 | m1(ABC)\*m2(AC)  0 | m1(ABC)\*m2(BC)  0 | m1(ABC)\*m2(A)  0 | m1(ABC)\*m2(B)  0 | m1(ABC)\*m2(C)  0 |
|  | m1(AB)\*m2(AB)  0 | m1(AC)\*m2(AC)  0 | m1(BC)\*m2(BC)  0 | m1(A)\*m2(ABC)  0 | m1(B)\*m2(ABC)  0 | m1(C)\*m2(ABC)  0 |
|  | m1(AB)\*m2(ABC)  0 | m1(AC)\*m2(ABC)  0 | m1(BC)\*m2(ABC)  0 | m1(A)\*m2(AB)  0 | m1(B)\*m2(AB)  0 | m1(C)\*m2(AC)  0 |
|  |  |  |  | m1(A)\*m2(AC)  0 | m1(B)\*m2(BC)  0 | m1(C)\*m2(BC)  0 |
|  |  |  |  | m1(A)\*m2(A)  0.21 | m1(B)\*m2(B)  0.04 | m1(C)\*m2(C)  0.05 |
|  |  |  |  | m1(AB)\*m2(A)  0 | m1(AB)\*m2(B)  0 | m1(AC)\*m2(C)  0 |
|  |  |  |  | m1(AC)\*m2(A)  0 | m1(BC)\*m2(B)  0 | m1(BC)\*m2(C)  0 |
| **0** | **0** | **0** | **0** | **0.21** | **0.04** | **0.05** |

Kết quả: 

|  |  |
| --- | --- |
| **ABC** | 0.0000 |
| **AB** | 0.0000 |
| **AC** | 0.0000 |
| **BC** | 0.7000 |
| **A** | 0.0000 |
| **B** | 0.1333 |
| **C** | 0.1667 |
| **** | 0.0000 |

Ví dụ 2: Cho Ω = {A,B,C}

m1(A) = 0.9, m1(B) = 0.05, m1(C) = 0.05

m2(A) = 0.1, m2(B) = 0.1, m2(C) = 0.8

Với Ω = {A,B,C}

|  |  |  |
| --- | --- | --- |
|  | **m1** | **m2** |
| **ABC** | 0.0000 | 0.0000 |
| **AB** | 0.0000 | 0.0000 |
| **AC** | 0.0000 | 0.0000 |
| **A** | 0.9000 | 0.1000 |
| **BC** | 0.0000 | 0.0000 |
| **B** | 0.0500 | 0.1000 |
| **C** | 0.0500 | 0.8000 |
| **Æ** | 0.0000 | 0.0000 |

Hệ số chuẩn hóa k:

|  |  |
| --- | --- |
| **BC=** |  |
| m1(ABC)\*m2() | 0 |
| m1(AB)\*m2(C) | 0 |
| m1(AB)\*m2() | 0 |
| m1(AC)\*m2(B) | 0 |
| m1(AC)\*m2() | 0 |
| m1(BC)\*m2(A) | 0 |
| m1(BC)\*m2() | 0 |
| m1(A)\*m2(B) | 0.14 |
| m1(A)\*m2(C) | 0.35 |
| m1(A)\*m2(BC) | 0 |
| m1(A)\*m2() | 0 |
| m1(B)\*m2(A) | 0.06 |
| m1(B)\*m2(C) | 0.1 |
| m1(B)\*m2(AC) | 0 |
| m1(B)\*m2() | 0 |
| m1(C)\*m2(A) | 0.03 |
| m1(C)\*m2(B) | 0.02 |
| m1(C)\*m2(AB) | 0 |
| m1(C)\*m2() | 0 |
| m1()\*m2(ABC) | 0 |
| m1()\*m2(AB) | 0 |
| m1()\*m2(AC) | 0 |
| m1()\*m2(BC) | 0 |
| m1()\*m2(A) | 0 |
| m1()\*m2(B) | 0 |
| m1()\*m2(C) | 0 |
| m1()\*m2() | 0 |
| **K** | **0.865** |

Hợp nhất 2 hàm khối lượng

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ABC** | **AB** | **AC** | **BC** | **A** | **B** | **C** |
| m1(ABC)\*m2(ABC)  0 | m1(ABC)\*m2(AB)  0 | m1(ABC)\*m2(AC)  0 | m1(ABC)\*m2(BC)  0 | m1(ABC)\*m2(A)  0 | m1(ABC)\*m2(B)  0 | m1(ABC)\*m2(C)  0 |
|  | m1(AB)\*m2(AB)  0 | m1(AC)\*m2(AC)  0 | m1(BC)\*m2(BC)  0 | m1(A)\*m2(ABC)  0 | m1(B)\*m2(ABC)  0 | m1(C)\*m2(ABC)  0 |
|  | m1(AB)\*m2(ABC)  0 | m1(AC)\*m2(ABC)  0 | m1(BC)\*m2(ABC)  0 | m1(A)\*m2(AB)  0 | m1(B)\*m2(AB)  0 | m1(C)\*m2(AC)  0 |
|  |  |  |  | m1(A)\*m2(AC)  0 | m1(B)\*m2(BC)  0 | m1(C)\*m2(BC)  0 |
|  |  |  |  | m1(A)\*m2(A)  0.09 | m1(B)\*m2(B)  0.005 | m1(C)\*m2(C)  0.04 |
|  |  |  |  | m1(AB)\*m2(A)  0 | m1(AB)\*m2(B)  0 | m1(AC)\*m2(C)  0 |
|  |  |  |  | m1(AC)\*m2(A)  0 | m1(BC)\*m2(B)  0 | m1(BC)\*m2(C)  0 |
| **0** | **0** | **0** | **0** | **0.09** | **0.005** | **0.04** |

Kết quả: 

|  |  |
| --- | --- |
| **ABC** | 0.0000 |
| **AB** | 0.0000 |
| **AC** | 0.0000 |
| **BC** | 0.6667 |
| **A** | 0.0000 |
| **B** | 0.0370 |
| **C** | 0.2963 |
| **** | 0.0000 |

1. New base basic assignment
2. Base basic function

Ví dụ 1: Cho Ω = {A,B,C}

m1(A) = 0.5, m1(B) = 0.45, m1(A, B, C) = 0.05

m2(A) = 0, m2(B) = 0.9, m2(C) = 0.1

Với Ω = {A,B,C} suy ra N = 3

mbase = = = 0.142857143

Ta có: mi\_base(Ai) = 

|  |  |  |  |
| --- | --- | --- | --- |
| **m1** | **m2** | **m1\_base** | **m2\_base** |
| 0.0500 | 0.0000 | 0.096428571 | 0.071428571 |
| 0.0000 | 0.0000 | 0.071428571 | 0.071428571 |
| 0.0000 | 0.0000 | 0.071428571 | 0.071428571 |
| 0.5000 | 0.0500 | 0.321428571 | 0.096428571 |
| 0.0000 | 0.0000 | 0.071428571 | 0.071428571 |
| 0.4500 | 0.9000 | 0.296428571 | 0.521428571 |
| 0.0000 | 0.1000 | 0.071428571 | 0.121428571 |
| 0.0000 | 0.0000 | 0.0000 | 0.0000 |

Hệ số chuẩn hóa k:

|  |  |
| --- | --- |
| **BC=** |  |
| m1(ABC)\*m2() | 0 |
| m1(AB)\*m2(C) | 0.008673469 |
| m1(AB)\*m2() | 0 |
| m1(AC)\*m2(B) | 0.037244898 |
| m1(AC)\*m2() | 0 |
| m1(BC)\*m2(A) | 0.006887755 |
| m1(BC)\*m2() | 0 |
| m1(A)\*m2(B) | 0.167602041 |
| m1(A)\*m2(C) | 0.039030612 |
| m1(A)\*m2(BC) | 0.022959184 |
| m1(A)\*m2() | 0 |
| m1(B)\*m2(A) | 0.028584184 |
| m1(B)\*m2(C) | 0.035994898 |
| m1(B)\*m2(AC) | 0.021173469 |
| m1(B)\*m2() | 0 |
| m1(C)\*m2(A) | 0.006887755 |
| m1(C)\*m2(B) | 0.037244898 |
| m1(C)\*m2(AB) | 0.005102041 |
| m1(C)\*m2() | 0 |
| m1()\*m2(ABC) | 0 |
| m1()\*m2(AB) | 0 |
| m1()\*m2(AC) | 0 |
| m1()\*m2(BC) | 0 |
| m1()\*m2(A) | 0 |
| m1()\*m2(B) | 0 |
| m1()\*m2(C) | 0 |
| m1()\*m2() | 0 |
| **K** | **0.417385204** |

Hợp nhất 2 hàm khối lượng

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ABC** | **AB** | **AC** | **BC** | **A** | **B** | **C** |
| m1(ABC)\*m2(ABC)  0.005102041 | m1(ABC)\*m2(AB)  0.005102041 | m1(ABC)\*m2(AC)  0.005102041 | m1(ABC)\*m2(BC)  0.005102041 | m1(ABC)\*m2(A)  0.006887755 | m1(ABC)\*m2(B)  0.006887755 | m1(ABC)\*m2(C)  0.037244898 |
|  | m1(AB)\*m2(AB)  0.005102041 | m1(AC)\*m2(AC)  0.005102041 | m1(BC)\*m2(BC)  0.005102041 | m1(A)\*m2(ABC)  0.037244898 | m1(B)\*m2(ABC)  0.006887755 | m1(C)\*m2(ABC)  0.006887755 |
|  | m1(AB)\*m2(ABC)  0.005102041 | m1(AC)\*m2(ABC)  0.005102041 | m1(BC)\*m2(ABC)  0.005102041 | m1(A)\*m2(AB)  0.037244898 | m1(B)\*m2(AB)  0.006887755 | m1(C)\*m2(AC)  0.006887755 |
|  |  |  |  | m1(A)\*m2(AC)  0.037244898 | m1(B)\*m2(BC)  0.006887755 | m1(C)\*m2(BC)  0.006887755 |
|  |  |  |  | m1(A)\*m2(A)  0.050280612 | m1(B)\*m2(B)  0.009298469 | m1(C)\*m2(C)  0.050280612 |
|  |  |  |  | m1(AB)\*m2(A)  0.006887755 | m1(AB)\*m2(B)  0.006887755 | m1(AC)\*m2(C)  0.037244898 |
|  |  |  |  | m1(AC)\*m2(A)  0.006887755 | m1(BC)\*m2(B)  0.006887755 | m1(BC)\*m2(C)  0.037244898 |
| **0.005102041** | **0.015306122** | **0.015306122** | **0.015306122** | **0.182678571** | **0.050625** | **0.182678571** |

Kết quả: 

|  |  |
| --- | --- |
| **ABC** | 0.0103 |
| **AB** | 0.0308 |
| **AC** | 0.0308 |
| **BC** | 0.3876 |
| **A** | 0.0308 |
| **B** | 0.1222 |
| **C** | 0.3876 |
| **** | 0 |

Ví dụ 2: Cho Ω = {A,B,C}

m1(A) = 0.2382, m1(B) = 0.2296, m1(C) = 0.2122, m1(A, B) = 0.0867, m1(A, C) = 0.074,

m1(B, C) = 0.085, m1(A,B,C) = 0.074

m2(A) = 0.2296, m2(B) = 0.1964, m2(C) = 0.208, m2(A,B) = 0.0866, m2(A,C) = 0.0954, m2(B,C) = 0.0971, m2(A,B,C) = 0.0866

Với Ω = {A,B,C} suy ra N = 3

mbase = = = 0.142857143

Ta có: mi\_base(Ai) = 

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **m1** | **m2** | **m1\_base** | **m2\_base** |
| ABC | 0.074 | 0.0866 | 0.108428571 | 0.114728571 |
| AB | 0.0867 | 0.0866 | 0.114778571 | 0.114728571 |
| AC | 0.074 | 0.0954 | 0.108428571 | 0.119128571 |
| A | 0.2382 | 0.2296 | 0.190528571 | 0.186228571 |
| BC | 0.085 | 0.0971 | 0.113928571 | 0.119978571 |
| B | 0.2296 | 0.1964 | 0.186228571 | 0.169628571 |
| C | 0.2122 | 0.208 | 0.177528571 | 0.175428571 |
| **** | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

Hệ số chuẩn hóa k:

|  |  |
| --- | --- |
|  |  |
| m1(ABC)\*m2() | 0 |
| m1(AB)\*m2(C) | 0.020135441 |
| m1(AB)\*m2() | 0 |
| m1(AC)\*m2(B) | 0.018392584 |
| m1(AC)\*m2() | 0 |
| m1(BC)\*m2(A) | 0.021216755 |
| m1(BC)\*m2() | 0 |
| m1(A)\*m2(B) | 0.032319089 |
| m1(A)\*m2(C) | 0.033424155 |
| m1(A)\*m2(BC) | 0.022859346 |
| m1(A)\*m2() | 0 |
| m1(B)\*m2(A) | 0.034681081 |
| m1(B)\*m2(C) | 0.032669812 |
| m1(B)\*m2(AC) | 0.022185144 |
| m1(B)\*m2() | 0 |
| m1(C)\*m2(A) | 0.033060892 |
| m1(C)\*m2(B) | 0.030113918 |
| m1(C)\*m2(AB) | 0.020367599 |
| m1(C)\*m2() | 0 |
| m1()\*m2(ABC) | 0 |
| m1()\*m2(AB) | 0 |
| m1()\*m2(AC) | 0 |
| m1()\*m2(BC) | 0 |
| m1()\*m2(A) | 0 |
| m1()\*m2(B) | 0 |
| m1()\*m2(C) | 0 |
| m1()\*m2() | 0 |
| **K** | **0.321425816** |

Hợp nhất 2 hàm khối lượng

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ABC** | **AB** | **AC** | **BC** | **A** | **B** | **C** |
| m1(ABC)\*m2(ABC)  0.012439855 | m1(ABC)\*m2(AB)  0.012439855 | m1(ABC)\*m2(AC)  0.012916941 | m1(ABC)\*m2(BC)  0.013009105 | m1(ABC)\*m2(A)  0.020192498 | m1(ABC)\*m2(B)  0.018392584 | m1(ABC)\*m2(C)  0.019021469 |
|  | m1(AB)\*m2(AB)  0.013168382 | m1(AC)\*m2(AC)  0.012916941 | m1(BC)\*m2(BC)  0.013668987 | m1(A)\*m2(ABC)  0.021859071 | m1(B)\*m2(ABC)  0.021365738 | m1(C)\*m2(ABC)  0.020367599 |
|  | m1(AB)\*m2(ABC)  0.013168382 | m1(AC)\*m2(ABC)  0.012439855 | m1(BC)\*m2(ABC)  0.013070862 | m1(A)\*m2(AB)  0.021859071 | m1(B)\*m2(AB)  0.021365738 | m1(C)\*m2(AC)  0.021148725 |
|  |  |  |  | m1(A)\*m2(AC)  0.022697397 | m1(B)\*m2(BC)  0.022343438 | m1(C)\*m2(BC)  0.021299624 |
|  |  |  |  | m1(A)\*m2(A)  0.035481864 | m1(B)\*m2(B)  0.031589687 | m1(C)\*m2(C)  0.031143584 |
|  |  |  |  | m1(AB)\*m2(A)  0.021375049 | m1(AB)\*m2(B)  0.019469725 | m1(AC)\*m2(C)  0.019021469 |
|  |  |  |  | m1(AC)\*m2(A)  0.020192498 | m1(BC)\*m2(B)  0.019325541 | m1(BC)\*m2(C)  0.019986327 |
| 0.012439855 | 0.038776618 | 0.038273737 | 0.039748955 | 0.163657447 | 0.15385245 | 0.151988798 |

Kết quả: 

|  |  |
| --- | --- |
| **ABC** | 0.01800 |
| **AB** | 0.05700 |
| **AC** | 0.05600 |
| **BC** | 0.24100 |
| **A** | 0.05900 |
| **B** | 0.22700 |
| **C** | 0.22400 |
| **** | 0.00000 |

1. **Chương 3: Cài đặt ứng dụng**
2. Giới thiệu dataset
   1. Tên bộ dữ liệu: winedata
   2. Tổng quan về bộ dữ liệu: Bộ dữ liệu bao gồm 178 dữ liệu, trong đó rượu được phân thành 3 loại: 1,2,3
   3. Thuộc tính của bộ dữ liệu: 13 thuộc tính:
      * Alcolhol: nồng độ cồn trong rượu
      * Malic acid: hàm lượng acid malic có trong rượu
      * Ash: tổng hàm lượng khoáng chất trong rượu
      * Alcalinity of ash: độ kiềm của rượu
      * Magnesium: hàm lượng magie trong rượu
      * Total phenols: tổng hàm lượng phenol trong rượu
      * Flavanoids. Nonflavanoid phenols: hàm lượng phenol không phải flavonoid có trong rượu
      * Proanthocyanins: hàm lượng proanthocyaninscos trong rượu
      * Color intensity: độ ẩm của màu sắc của rượu
      * Hue: độ bão hòa của màu đỏ và màu xanh lá cây trong rượu
      * OD280/OD315 of diluted wines: tỉ lệ giữa độ hấp thụ ánh sáng ở hai bước sóng khác nhau
      * Proline: hàm lượng amino acid có trong rượu
3. Cài đặt các thuật toán
4. Kết quả cài đặt
5. **Kết luận và kiến nghị**
6. **Tài liệu tham khảo và phụ lục**

**Phụ lục**

[**I.** **Mở đầu** 1](#_Toc133878486)

[**II.** **Nội dung và kết quả nghiên cứu** 1](#_Toc133878487)

[**A.** **Chương 1: Cơ sở lý thuyết** 1](#_Toc133878488)

[1. Lý thuyết bằng chứng Demster-Shafer 1](#_Toc133878489)

[2. Một phương pháp quản lý xung đột hiệu quả 2](#_Toc133878490)

[3. 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 12](#_Toc133878491)

[4. Phương pháp áp dụng lý thuyết chứng cứ vào việc hợp nhất dữ liệu mâu thuẫn 41](#_Toc133878492)

[**B.** **Chương 2: Các luật hợp nhất tri thức** 62](#_Toc133878493)

[1. Dempster’s combination rule 62](#_Toc133878494)

[2. New base basic assignment 62](#_Toc133878495)

[3. Proposed method 62](#_Toc133878496)

[**C.** **Chương 3: Cài đặt ứng dụng** 62](#_Toc133878497)

[1. Giới thiệu dataset 62](#_Toc133878498)

[2. Cài đặt các thuật toán 63](#_Toc133878499)

[3. Kết quả cài đặt 63](#_Toc133878500)

[**III.** **Kết luận và kiến nghị** 63](#_Toc133878501)

[**IV.** **Tài liệu tham khảo và phụ lục** 63](#_Toc133878502)