



Universitatea Tehnică “Gheorghe Asachi” Iași

Facultatea de Automatică și Calculatoare Iași

Disciplina: Inteligență Artificială

Proiect - Regula Dempster–Shafer

Studenți:

Petcu Andrei-Dănuț,

Ocu Eusebiu

Grupa: 1407A

Anul: IV

Profesor coordonator:

Mircea Hulea

An universitar 2025–2026

1. Introducere

În cadrul disciplinei Inteligență Artificială, un aspect important îl reprezintă luarea deciziilor în condiții de incertitudine. În multe aplicații reale, informațiile disponibile sunt incomplete, imprecise sau provin din surse diferite care pot fi contradictorii. Metodele clasice bazate pe probabilități presupun, de regulă, existența unor informații complete și precise, ceea ce nu este întotdeauna posibil în practică.

Teoria evidențelor, cunoscută sub numele de teoria Dempster–Shafer, a fost dezvoltată ca o alternativă la teoria probabilităților clasice, având ca scop tratarea incertitudinii și a lipsei de informații. Această teorie permite reprezentarea explicită a necunoașterii și oferă un mecanism formal pentru combinarea mai multor surse de informație.

Un avantaj important al teoriei Dempster–Shafer este faptul că permite alocarea unei părți din informație către mulțimi de ipoteze, nu doar către ipoteze individuale. Astfel, sistemul poate exprima situații de tipul „nu sunt sigur între două opțiuni”, lucru imposibil în cadrul probabilităților clasice.

Scopul acestui proiect este implementarea regulii Dempster–Shafer în limbajul de programare Java și aplicarea acesteia în două probleme reale distincte: diagnosticul medical și evaluarea riscului de accident rutier. Prin aceste aplicații se urmărește evidențierea modului în care teoria evidențelor poate fi utilizată în practică pentru a obține decizii mai bine fundamentate.

2. Descrierea generală a problemei de agregare a evidențelor

Agregarea evidențelor reprezintă procesul prin care mai multe surse de informație sunt combinate pentru a obține o concluzie unică. În viața reală, rareori o singură sursă este suficientă pentru a lua o decizie corectă. De cele mai multe ori, informațiile provin din surse diferite, fiecare având un anumit grad de încredere.

Problema apare atunci când aceste surse nu sunt complet de acord sau când nu furnizează informații precise. De exemplu, în domeniul medical, simptomele pacientului pot sugera o anumită boală, în timp ce rezultatele analizelor pot indica o altă afecțiune. În mod similar, în domeniul siguranței rutiere, condițiile meteo pot indica un risc ridicat, în timp ce traficul este redus.

Teoria Dempster–Shafer oferă un cadru matematic prin care aceste evidențe pot fi combinate, ținând cont atât de informațiile furnizate, cât și de gradul de conflict dintre ele. Rezultatul final

nu este o singură valoare de probabilitate, ci un set de valori care exprimă gradul de credință asociat fiecărei ipoteze.

3. Problema 1 – Diagnostic medical

Diagnosticul medical este un domeniu în care incertitudinea joacă un rol major. Medicii iau decizii pe baza simptomelor raportate de pacient, a rezultatelor testelor medicale și a experienței clinice. Aceste surse pot furniza informații incomplete sau chiar contradictorii.

În această aplicație se consideră un set simplificat de boli posibile: gripă, COVID-19 și alte afecțiuni. Fiecare dintre aceste ipoteze reprezintă o posibilă stare a pacientului. Sursele de evidență utilizate sunt simptomele pacientului, testul rapid și evaluarea clinică realizată de medic.

Pentru fiecare sursă se definește o funcție de masă care exprimă gradul de credință asociat fiecărei ipoteze sau combinații de ipoteze. De exemplu, simptomele pot indica o probabilitate mai mare pentru gripă, dar pot lăsa loc și pentru incertitudine între gripă și COVID-19.

Prin aplicarea regulii Dempster–Shafer, aceste evidențe sunt combinate progresiv, obținându-se o distribuție finală a credinței. Această distribuție permite identificarea afecțiunii cu cel mai mare grad de susținere, dar și evaluarea nivelului de incertitudine rămas.

4. Problema 2 – Evaluarea riscului de accident rutier

Evaluarea riscului de accident rutier este o problemă importantă în domeniul siguranței transporturilor. Accidentele rutiere sunt influențate de numeroși factori, precum condițiile meteo, densitatea traficului și starea infrastructurii.

În această aplicație, riscul rutier este clasificat în trei niveluri: risc scăzut, risc mediu și risc ridicat. Sursele de evidență utilizate sunt condițiile meteo, nivelul de trafic și starea drumului.

Fiecare sursă furnizează o funcție de masă care reflectă gradul de influență asupra riscului de accident. De exemplu, condițiile meteo nefavorabile pot indica un risc ridicat, în timp ce starea bună a drumului poate reduce acest risc.

Prin combinarea acestor evidențe utilizând regula Dempster–Shafer, se obține o evaluare finală a riscului rutier, care poate fi utilizată pentru luarea unor decizii preventive.

5. Aspecte teoretice privind algoritmul Dempster–Shafer

Teoria Dempster–Shafer este o extensie a teoriei probabilităților clasice și este cunoscută sub numele de teoria evidențelor. Aceasta se bazează pe conceptul de cadru de discernământ, notat Θ , care reprezintă mulțimea tuturor ipotezelor posibile.

O funcție de masă, notată m , atribuie un grad de credință fiecărei submulțimi a lui Θ . Spre deosebire de probabilități, masa de credință poate fi atribuită și unor submulțimi care conțin mai multe ipoteze, reflectând astfel incertitudinea.

Regula de combinare Dempster–Shafer permite agregarea a două funcții de masă independente. Aceasta se bazează pe calcularea intersecțiilor dintre submulțimile asociate evidențelor și pe redistribuirea masei de conflict.

Conflictul apare atunci când două surse susțin ipoteze incompatibile. În regula standard Dempster–Shafer, conflictul este eliminat prin normalizare, iar masa rămasă este redistribuită proporțional.

6. Modalitatea de rezolvare și implementare

Implementarea a fost realizată în limbajul Java, utilizând structuri de date simple, precum EnumSet și Map. Acestea permit reprezentarea eficientă a submulțimilor ipotezelor și a valorilor asociate acestora.

Algoritmul de combinare a evidențelor parcurge toate combinațiile posibile dintre submulțimile a două surse și calculează intersecția acestora. În cazul în care intersecția este vidă, contribuția este considerată conflict.

După calcularea masei totale de conflict, valorile obținute sunt normalizate, conform regulii Dempster–Shafer. Rezultatul final este o nouă funcție de masă care reflectă aggregarea celor două surse.

Aceeași implementare a algoritmului a fost utilizată pentru ambele aplicații, demonstrând astfel caracterul general al metodei.

7. Rezultate experimentale

7.1 Metodologia testării

Pentru evaluarea funcționării aplicațiilor dezvoltate, au fost realizate mai multe teste folosind valori diferite ale funcțiilor de masă asociate fiecărei surse de evidență. Scopul acestor teste a fost observarea comportamentului regulii Dempster–Shafer în situații variate, precum evidențe convergente, evidențe conflictuale și situații cu un nivel ridicat de incertitudine.

Testele au fost realizate prin rularea aplicațiilor în mediul de dezvoltare IntelliJ IDEA, iar rezultatele au fost afișate în consola aplicației. Pentru fiecare scenariu au fost analizate distribuțiile finale ale maselor de credință, precum și evoluția acestora după fiecare etapă de agregare.

7.2 Rezultate pentru problema de diagnostic medical

7.2.1 Implementarea scenariului de test – definirea evidențelor

În figura următoare este prezentat fragmentul de cod utilizat pentru definirea evidențelor. Se observă faptul că, pe lângă ipoteze individuale (ex. {GRIPA}), se folosesc și mulțimi care conțin mai multe ipoteze (ex: {GRIPA, COVID}) sau chiar mulțimea THETA (care înseamnă „necunoaștere / nu sunt sigur”).

```
public static void main(String[] args) { ➜ Andrei Petcu

    EnumSet<Hypothesis> THETA = EnumSet.allOf(Hypothesis.class);

    Map<EnumSet<Hypothesis>, Double> symptoms = new HashMap<>();
    symptoms.put(EnumSet.of(Hypothesis.GRIPA), 0.6);
    symptoms.put(EnumSet.of(Hypothesis.COVID), 0.2);
    symptoms.put(EnumSet.of(Hypothesis.GRIPA, Hypothesis.COVID), 0.2);

    Map<EnumSet<Hypothesis>, Double> test = new HashMap<>();
    test.put(EnumSet.of(Hypothesis.COVID), 0.7);
    test.put(EnumSet.of(Hypothesis.GRIPA), 0.1);
    test.put(THETA, 0.2);

    Map<EnumSet<Hypothesis>, Double> doctor = new HashMap<>();
    doctor.put(EnumSet.of(Hypothesis.GRIPA), 0.4);
    doctor.put(EnumSet.of(Hypothesis.COVID), 0.3);
    doctor.put(THETA, 0.3);
```

Interpretarea valorilor alese în acest scenariu este următoarea:

- Pentru evidența symptoms (simptome), s-a ales o masă mai mare pentru ipoteza {GRIPA}, deoarece simptomele pot sugera gripă. O parte din masă este alocată mulțimii {GRIPA, COVID}, ceea ce exprimă faptul că, doar din simptome, există incertitudine între cele două afecțiuni (unele simptome sunt similare).
- Pentru evidența test (test rapid), masa de credință este dominantă pe {COVID}, reflectând un rezultat care sugerează COVID. Totuși, există și o componentă de necunoaștere THETA, deoarece testul poate avea erori sau poate fi neconcludent.
- Pentru evidența doctor (evaluarea medicului), există o distribuție mai echilibrată și o masă semnificativă pe THETA, deoarece medicul poate considera că informațiile sunt încă insuficiente pentru o concluzie 100% sigură.

7.2.2 Scenariul 1 – Agregarea simptome + test rapid

După apelul metodei DempsterShafer.combine(symptoms, test), se obține o nouă funcție de masă (rezultatul first) care sintetizează informațiile din cele două surse. Rezultatul este afișat în consolă, iar în figura următoare este prezentată captura obținută în urma rulării.

```
Agregare 1: simptome + test
[GRIPA] -> 0.35714285714285715
[COVID] -> 0.5714285714285713
[GRIPA, COVID] -> 0.07142857142857144
```

7.2.3 Scenariul 2 – Agregarea rezultatului cu evaluarea medicului

Prin acest pas se urmărește integrarea unei surse de informație care conține, de regulă, mai multă incertitudine (medicul nu are încă toate datele), dar poate adăuga un echilibru în cazurile în care primele două surse sunt conflictuale. Rezultatul final este stocat în variabila second și este afișat în consolă.

```
Agregare 2: rezultat + evaluare medic
[GRIPA] -> 0.4193548387096774
[COVID] -> 0.5483870967741933
[GRIPA, COVID] -> 0.03225806451612903
```

7.2.4 Observații privind alegerea valorilor și interpretarea rezultatelor

Valorile numerice utilizate pentru funcțiile de masă reprezintă o modelare simplificată și au fost alese manual pentru a simula scenarii diferite. Chiar dacă aceste valori nu provin dintr-un set medical real, ele reflectă situații plauzibile din punct de vedere logic: de exemplu, un test rapid poate indica puternic o afecțiune, dar poate lăsa o parte din masă pe necunoaștere (THETA) din cauza limitărilor testului.

În mod ideal, într-o aplicație reală, aceste valori ar putea fi învățate din date (statistici medicale, acuratețea testelor etc.). Totuși, pentru scopul proiectului, este important să se demonstreze implementarea corectă a regulii Dempster–Shafer și modul în care agregarea evidențelor produce un rezultat final coerent.

7.3 Rezultate pentru problema de evaluare a riscului de accident rutier

7.3.1 Implementarea scenariului de test – definirea evidențelor

În aplicația de evaluare a riscului de accident rutier, fiecare sursă de informație este reprezentată sub formă unei funcții de masă (BBA) implementată în Java printr-un Map<EnumSet<Risk>, Double>. Cheia map-ului reprezintă o submulțime din cadrul de discernământ Θ , iar valoarea reprezintă masa de credință alocată acelei submulțimi.

Cadrul de discernământ pentru această problemă este format din trei niveluri de risc:

- RISC_SCAZUT
- RISC_MEDIU
- RISC RIDICAT

```
public static void main(String[] args) {

    EnumSet<Risk> THETA = EnumSet.allOf(Risk.class);

    Map<EnumSet<Risk>, Double> weather = new HashMap<>();
    weather.put(EnumSet.of(Risk.RISC RIDICAT), 0.55);
    weather.put(EnumSet.of(Risk.RISC_MEDIU), 0.25);
    weather.put(THETA, 0.20);

    Map<EnumSet<Risk>, Double> traffic = new HashMap<>();
    traffic.put(EnumSet.of(Risk.RISC RIDICAT), 0.40);
    traffic.put(EnumSet.of(Risk.RISC_MEDIU), 0.40);
    traffic.put(THETA, 0.20);

    Map<EnumSet<Risk>, Double> road = new HashMap<>();
    road.put(EnumSet.of(Risk.RISC_MEDIU), 0.30);
    road.put(EnumSet.of(Risk.RISC_SCAZUT), 0.20);
    road.put(EnumSet.of(Risk.RISC RIDICAT), 0.20);
    road.put(THETA, 0.30);
```

7.3.2 Afişarea evidenţelor iniţiale

Programul afişează mai întâi evidenţele iniţiale, pentru ca utilizatorul să poată observa valorile de intrare ale funcţiilor de masă. Această etapă este utilă şi în documentaţie, deoarece se poate vedea clar cum sunt definite sursele înainte de agregare.

```
==== Evidenta 1 (Vreme) ====
[RISC RIDICAT] -> 0.55
[RISC_SCAZUT, RISC_MEDIU, RISC RIDICAT] -> 0.2
[RISC_MEDIU] -> 0.25
```

```
==== Evidenta 2 (Trafic) ====
[RISC RIDICAT] -> 0.4
[RISC_SCAZUT, RISC_MEDIU, RISC RIDICAT] -> 0.2
[RISC_MEDIU] -> 0.4
```

```
==== Evidenta 3 (Drum) ====
[RISC RIDICAT] -> 0.2
[RISC_SCAZUT, RISC_MEDIU, RISC RIDICAT] -> 0.3
[RISC_MEDIU] -> 0.3
[RISC_SCAZUT] -> 0.2
```

7.3.3 Scenariul 1 – Agregarea Vreme \oplus Trafic

Apelul DempsterShafer.combine(weather, traffic) produce o nouă funcție de masă (combined12) care reprezintă informația combinată din cele două surse. Rezultatul este afișat în consolă.

```
== Agregare 1: Vreme  $\oplus$  Trafic ==
[RISC RIDICAT] -> 0.6029411764705884
[RISC_SCAZUT, RISC_MEDIU, RISC RIDICAT] -> 0.058823529411764726
[RISC_MEDIU] -> 0.3382352941176471
```

7.3.4 Scenariul 2 – Agregarea rezultatului cu starea drumului

A doua etapă combină rezultatul primei agregări cu a treia sursă de evidență (starea drumului). În aplicație, această etapă este evidențiată prin mesajul:

```
== Agregare 2: (Vreme  $\oplus$  Trafic)  $\oplus$  Drum ==
[RISC RIDICAT] -> 0.5561357702349873
[RISC_SCAZUT, RISC_MEDIU, RISC RIDICAT] -> 0.03133159268929506
[RISC_MEDIU] -> 0.39164490861618806
[RISC_SCAZUT] -> 0.02088772845953004
```

7.3.5 Determinarea concluziei finale

După calcularea distribuției finale, aplicația determină concluzia finală într-o manieră simplă: alege submulțimea cu valoarea numerică cea mai mare (cea mai mare masă de credință). Această logică este implementată printr-o parcurgere a map-ului combined123.

```
Concluzie: cel mai probabil = [RISC RIDICAT] (valoare 0.5561357702349873)
```

7.3.6 Observații generale asupra rezultatului și asupra valorilor utilizate

Valorile numerice utilizate pentru masele de credință au fost alese manual pentru a simula un scenariu de trafic realist. De exemplu, vremea are o influență considerabilă asupra riscului, motiv pentru care în evidență weather masa dominantă este acordată riscului ridicat. Traficul are o distribuție aproape egală între risc mediu și ridicat, reflectând faptul că nu toate situațiile de trafic intens duc la accidente, dar cresc probabilitatea acestora. Starea drumului a fost modelată astfel încât să includă și posibilitatea de reducere a riscului (prin prezența unui risc scăzut).

Într-un sistem real, aceste mase ar putea proveni din date istorice (statistici de accidente, informații de la senzori, rapoarte meteo etc.). Totuși, pentru acest proiect, obiectivul principal este implementarea corectă a regulii Dempster–Shafer și evidențierea modului în care aceasta se poate aplica într-un domeniu cu informații incerte.

8. Concluzii

În urma realizării acestui proiect, s-a demonstrat că teoria Dempster–Shafer este o metodă eficientă pentru agregarea informațiilor incerte. Implementările realizate confirmă flexibilitatea metodei și aplicabilitatea acesteia în domenii diferite.

Comparativ cu metodele probabilistice clasice, teoria evidențelor oferă un avantaj important prin posibilitatea reprezentării explicite a necunoașterii. Utilizarea mai multor surse de evidență conduce la decizii mai robuste și mai bine fundamentate.

9. Bibliografie

- [1] Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976
- [2] Curs Inteligență Artificială – Metode de raționament cu informații incomplete
- [3] Russell, S., Norvig, P., *Artificial Intelligence: A Modern Approach*

10. Contribuția membrilor echipei

Petcu Andrei-Dănuț:

- implementarea aplicației de diagnostic medical;
- definirea evidențelor și scenariilor;
- redactarea părții aferente diagnosticului medical.

Ocu Eusebiu:

- implementarea aplicației de evaluare a riscului rutier;
- definirea evidențelor specifice traficului;
- redactarea părții aferente evaluării riscului rutier.