# Effektívny fuzzy klasifikátor s možnosťou výberu na základe fuzzy entropie

Abstract

Tento papier rozpráva o effektívnom fuzzy klasifikátore s možnosťou výberu založenom na meraní fuzzy entropie.

Fuzzy entropia je použitá na vyhodnotenie informácie o distribúcii vzorov v priestore vzorov. S touto informáciou vieme rozdeliť priestor vzorov na disjunktné rozhodovacie regióny pre rozoznávanie vzorov. Vďaka tomu, že rozhodovacie regióny sú disjunktné, aj komplexnosť, aj výpočtová náročnosť je zredukovaná, a tým pádom aj čas trénovania a klasifikácie je extrémne krátky. Hoci rozhodovacie regióny sú rozdelené do disjunktných podpriestorov, môžeme dosiahnuť kvalitnú klasifikáciu vďaka tomu, že podpriestori boli správne stanovené naším navrhovaným merania fuzzy entropie. Okrem toho môžeme skúmať ďalšie využitie fuzzy entropie na vybraté prvky. Procedúra výberu prvkov nielenže znižuje dimenziu problému, ale aj redukuje šum, zbytočné a nedôležité prvky. Nakoniec aplikujeme navrhnutý klasifikátor do databázi Iris a Wiscinsinskej databázi rakoviny prsníka na vyhodnotenie vúkonu klasifikátora. Obe výsledky ukazujú, že navrhnutý klasifikátor môže dobre fungovať s aplikáciou na rozlišovanie vzorov.

1. Úvod

Cieľom tejto práce je navrhnúť efektívny fuzzy klasifikátor s možnosťou výberu vlastnosti/funkcie/prvku. Meranie fuzzy entropie je využité na rozdelenie vstupného priestoru vlastností/prvkov do rozhodovacích oblastí a na výber súvisiacich/vhodných vlastností na dobré rozdelenie úlohy klasifikácie. Pre problém klasifikácie, metódy na uzatvorenie rozhodových regiónov a rozmery/dimenzie problému majú silný vpliv na rýchlosť a výkon klasifikácie. V tomto článku sú disjunktné fuzzy rozhodovacie regióny zovšeobecnené. Hoci rozhodovacie regióny sa neprekrývajú, môžme stále získať správne ohraničenia z fuzzy rozhodovacích regiónov na dosiahnutie dobrého výkonu klasifikácie. Okrem toho, obe výpočtová náročnosť a zložitosť sú zredukované, takže klasifikácia môže byť extrémne rýchla. Navyše navrhnutá metóda výberu vlastnosti na základe fuzzy entropie zvyšuje rýchlosť klasifikácie vynechaním šumivých, opakujúcich sa a nedôležitých prvkov/vlastností.

Dve dôležité problémy pri vývojo klasifikátora sú redukcia času a zvýšenie miery klasifikácie [1]. Rozličné spôsoby boli navrhnuté na vývoj fuzzy klasifikátorov [1] - [3]. Jednoduchou metódou na skonštruovanie klasifikátora je použitie neurónovej siete, ale na druhú stranu trvá príliš veľa času na natrénovanie takejto neurónovej siete [1]. Taktiež výsledky sa ťažko analyzujú, takže je náročné zlepšiť výkon natrénovanej siete. Na vyriešenie tohto problému vedci navrhli viacero spôsobov založených na analýze fuzzy regiónov pre skonštruovanie fuzzi klasifikátora [1], [2], [4]-[8]. Vo všeobecnosti, “tvary” fuzzy regiónov môžeme zaradiť do dvoch druhov:

1. Hyperbolické regióny [2], [4], [6], ktorých hranice sú paralélne so vstupnými osami (priestoru);
2. Elipsoidné regióny [1] a [9];
3. Mnohostenné regióny, ktorých hranice sú vyjadrené lineárnou kombináciou vstupných parametrov

Klasifikátor používajúci hyperbolické regióny je jednoduchší a menej časovo náročný než tie, čo používajú elipsoidné čo mnohostenné regióny [1].

Naša navrhovaná metóda je založená na hyperbolických regiónoch, ale fuzzy regióny sa v nich neprekrývajú ako v [2]. Používame meranie fuzzy entropie namiesto fuzzy pravidiel, na odzrkadlenie aktuálnej distribúcie klasifikačných vzorov pomocou rátania fuzzy entropie pre každý rozmer/dimenziu. Rozhodovacie regióny môžu byť dolaďované automaticky podla merania fuzzy entropie. Taktiež, kvôli schopnosti zdetekovať aktuálnu distribúciu/rozmiestnenie vzorov, generované rozhodovacie regióny môžu byť efektívne zredukované.

Vzhľadom k tomu, že najnovšie klasifikačné systémy majú za ciel vyrovnať sa s väčšími a komplexnejšími úlohami, problém výberu vlastnosti naberá na dôležitosti [10], [11]. Vo všeobecnosti sú dva typy výberu vlastnosti/funkcie: 1) výberom príslušnej podmnožiny vlastností/funkcií z originálnej množiny vlastností/funkcií a 2) transformáciou alebo kombináciou originálnych funkcií do nových, čo je uvedená ako *feature extraction*(extrakcia funkcií/vlastností) [12], [13].

Extrakcia funkcie zvyšuje schopnosť nájsť nové funkcie na popis priestoru vzorov [13]. Je to obzvlášť užitočné keď originálne vlastnosti/funkcie neadekvátne rozdelujú triedy, napr. nemôžeme vybrať optimálnu podmnožinu vlastností/funkcií priamo zo zadanej množiny. Ale extrakčná metóda vyžaduje viac námahy a transformované vlastnosti/funkcie môžu stratiť fyzický význam originálnych vlastností/funkcií. Oproti tomu, prístup s výberom vlastnosti zachováva fyzický význam vybratých vlastností/funkcií. Pre systémy založené na pravidlách je uchovanie fyzického významu vlastnosti/funkcie dôležité a nepostrádateľné [13].

Výber vlastnosti/funkcie bol študovaný po mnoho rokoch a viacero prístupov bolo navrhnutých [10], [11], [14], [15]. Viaceré z nich sú založené na neurónových sietiach [16]–[20] napriek tomu, že tento spôsob je časovo náročný. Okrem toho vybrané funkcie/vlastnosti nemusia by interpretovateľné a nemusia byť vhodné pre systémy klasifikácie. Podobne, prístup zaločený na genetickom algoritme pre výber vlastnosti/funkcie má taktiež problém s časovou náročnosťou [21]–[24]. Na vyhnutie sa tohto problému preto navrhujeme efektívnu metódu pre výber relevantných vlastností/funkcií založené na fuzzy entropie. (to znie ako z teleshopingu, [poznámka redaktora]) Boli navrhnuté rozličné definície fuzzy entropie [25], [26]. (oh, really? Thanks a lot. A to som čakal, že konečne povie, čo to fuzzy entropia je, [poznámka redaktora]) V podstate, dobre zadefinované meranie fuzzy entropie musí spĺňať 4 Luca-Termini axiómy [25], bližšie popísané v Sekcii II. Naša navrhovaná fuzzy entropia je odvodená od Shannonovej entropie [27] a spĺňa tie 4 axiómy.

Článok je nasledovne organizovaný: Sekcia II sa vzťahuje na Shannonovu entropiu a prezentuje náš návrh entropie. Sekcia III popisuje operácie fuzzy klasifikátora, vrátane určenia číselných intervalov, stredu a šírky každého intervalu a určenie častí rozhodovacích regiónov. Sekcia IV ukazuje proces výberu vlastnosti, a v Sekcii V sú zobrazené niektoé výsledky experimentov. Diskusia a záver sú v Sekcii VI.

1. Meranie Entropie

Entropia je meraná množstvom neistoty výsledky náhodného experimentu, alebo equivalene, meraním informácií keď sa pozoruje výsledok. Tento koncept bol zadefinovaný rôznymi spôsobmi [25]–[30] a zovšeobecnený v rozličných aplikovaných oblastiach, ako napríklad teória komunikácie, matematiky, štatistickej termodynamike a ekonómii [31]–[33]. Z pomedzi týchto rozličných definícií, Shannon prispel k najširšej a najfundamentálnejšej definícii entropie v informačnej teórii. V nasledujúcom texte najprv uvedieme Shannonovu entropiu a potom popíšeme štyry Luca-Termini axiómy [25], ktoré dobre-definovaná entropia musí spĺňať. Nakoniec navrhneme meranie fuzzy entropie, ktoré je rozšírenie Shannonovej definície.

*A. Shannonova Entropia*

Za entropiu možno považovať meranie neistoty náhodnej premennej *X*. Nech *X* je náhodná spočítateľná premenná s konečnou *N*-znakovou abecedou danou .

Ak výsledok sa vyskytuje s pravdepodobnosťou , tak potom množstvo informácie spojené so známim výskytom výstupu je definované ako:

(1.)

To znamená, že pre diskrétne zdroje, informácie získané výberom symbolu súbitové. V priemere, symbol bude vybratý -krát z celkového počtu výberov, takže priemerné množstvo informácie získanej z zdrojových výsledkou je:

(2.)

Deliac (2.) číslom získame priemerné množstvo informácie na symbol výstubu zdroja. To je známe ako priemerná informácia, neistota, alebo entropia definovaná nasledovne.

Definícia 1: Entropia H(X) náhodnej diskrétnej premennej X je definovaná ako [27]

(3.)

Alebo

(4.)

Kde znamená .

Všimnite si, že entropia je funkcia distribúcie . Nezáleží na skutočných hodnotách náhodnej premennej , ale iba na pravdepodobnostiach. Preto entropiu možno zasísať ako .

*B. Luca-Termini Axómy pre Fuzzy Entropiu*

Kosko [25] navrhuje, aby meranie dobre definovanej fuzzy entropie musí spĺňať štyry Luca-Termini exiómy. Patria medzi ne nasledujúce:

1. iff , kde nie je fuzzy množina a je množina všetkých podmnožín množiny .
2. iff pre všetky , kde znamená stupeň zastúpenia vo fuzzy množine .
3. ak je menej fuzzy než , napr. ak keď a keď , kde aj sú fuzzy množiny.
4. .

*C. Fuzzy Entropia*

Náš návrh fuzzy entropie založený na Shannonovej entropie je nasledne definovaný:

*Defininícia 2: Fuzzy Entropia intervalu pre každý rozmer prvku:*

1. Nech je univerzálna množina prvkov v priestore, kde .
2. Nech je fuzzy množina definovaná na intervale priestoru, ktorý obshuje elementov (k<n). Namapovanie stupňa príslušnosti prvku s fuzzy množinou sa označuje ako .
3. Nech reprezentujú tried v ktorých je prvkov rozdelených.
4. Nech je množina prvkov triedy na univerzálnej množine . Je to podmnožina univerzálnej množiny .
5. Stupeň priradenia prvkov fuzzy množiny do triedy v intervale, kde je defonovaný ako:  
    (5)
6. Fuzzy Entropia prvkov triedy v intervale je definované ako:  
    (6) .
7. Fuzzy Entropia na univerzálnej množine pre prvky v intervale je definovaná ako:  
    (7)

V definícii (6) je fuzzy entropia ako nepravdepodobnostná entropia. Preto môžme zadefinovať nový pojem “stupeň priradenia” pre . Základná vlastnosť navrhnutej fuzzy entropie je podobná ako Shannonova entropia a spĺňa štyry Luca-Termini axiómy, ale ich spôsoby merania informácie sú rôzne. Pravdepodobnosť Shannonovej entropie je meraná cez výskyt prvku. Oproti tomu, stupeň priradenia vo fuzzy entropii je merané členstvom hodnôt vyskytujúcich sa prvkov. Okrem toho, fuzzy entropia rozhodovacích regiónoch môže byť získaná cez súčet fuzzy entropie jednotlivých intervalov v každej dimenzii vlastností. Nakoniec zobrazíme rozdiel medzi navrhovanou a Shannonovou entropiou meraním informácie v dvoch rozdeleniach, ako je zobrazené na Obrázku 1.

Keďže pravdepodobnosti týchto dvoch distribúcií v každom intervale vzorkového priestoru [0,1] na Obr. 1(a) aj (b) sú rovnaké, one distribúcie majú rovnakú Shannonovu entropiu. Napríklad výpočet pre Shannonovej entropie v intervale je preukázané nasledovne:

Pravdepodobnosť výskytu “\*” je , pravdepodobnosť je .

Shannonova entropia je:

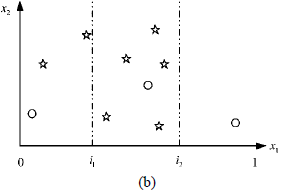
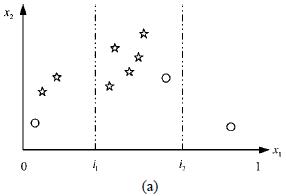
.

Ako bolo uvedené vyššie, stupeň zhody vo fuzzy entropii je založený na mapovaní členstva hodnôt prvkov. Predpokladajme, že začneme priraďovaním troch členských funkcií s prekrývajúcimi sa regiónmi vo vzorkovom priestore [0,1], ako je zobrazené na Obr. 2, pre vzorkový priestor na Obr. 1. Hodnotu členskej funkcie možno považovať za mieru, do ktorej vzorka patrí v špecifikovanom vzorkovom priestore.

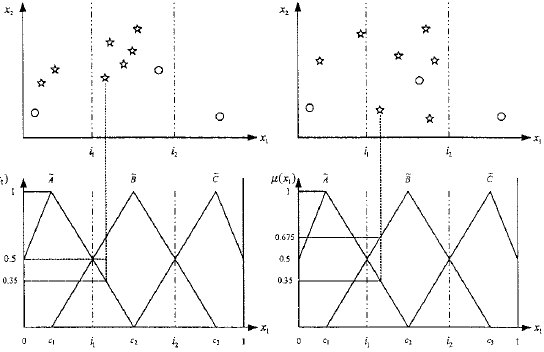
Fuzzy entropia intervalu pre dve rozdelenia ukázané na Obr. 2 sú nasledovné:

*Rozdelenie (a):*

1. Z zodpovedajúcej členskej funkcie , celková miera zastúpenia “\*” je 0,35+0,3+0,1+0,0+0,0 = 0,75.  
   Celková miera zastúpenia je 0,0.



Obr. 1. Dva príklady distribúcie vzorov s dvoma vlastnosťami a triedami (\* a O sú prvky triedy 1 a triedy 2, resp.)



Obr. 2. Dva príklady distribúcie vzorov s korenšpondujúcimi členskými funkciami. (sú stredy troch triangulárnych fuzzy množín)

Stupeň zastúpenia “\*” je .

Stupeň zastúpenia “O” je .

Fuzzy entropia je:

Preto fuzzy entropia pre prvky v intervalev rozmere je:

2) Zo zodpovedajúcej funkcie členstva , celkový stupeň členstva “\*” je .  
Celkové zastúpenie “O” je 0.7.  
Stupeň zastúpenia “\*” je .  
Stupeň zastúpenia “O” je .  
Fuzzy entropia je:

.

Fuzzy entropia pre prvky v intervale rozmeru je:

.

3) Zo zodpovedajúcej funkcie členstva , celkový stupeň členstva “\*” je .  
Celkový stupeň začlenenia “O” je 0,3.  
Stupeň zastúpenia “\*” je .  
Stupeň zastúpenia “\*” je .  
Fuzzy entropia je

.

Fuzzy entropia pre vzorky intervalu v dimenzii je:

.

Nakoniec získame fuzzy entropiu zosumovaním všetkých fuzzy entropií nasledovne:

.

Rozdelenie (b):

S použitím rovnakej metódy počítania ako v rozdelení (a), môžeme získať celú fuzzy entropiu

(b) v intervale nasledovne:

.

Ako je uvedené vyššie, fuzzy entropia rozdelenia (a) je menšia ako rozdelenia (b). Tento výsledok znamená, že výskyt “\*” v rozdelení (a) je viac usporiadaný ako v (b). Pozerajúc sa na aktuálne rozdelenia, výsledok dáva zmysel. Oproti tomu, používajúc Shannonovu entropiu, hodnoty pre tieto dve rozdelenia sú identické pre interval .

Z obrázkou uvedených vyššie môžeme vydieť, že navrhnutá fuzzy entropia je lepšie schopná rozlíšiť skutočné rozloženie vzorov. Použitím členskej funkcie pre meranie stupňa príšlusnosti, hodnota entropie obsahuje nie len počet vzorov, ale aj berie do úvahy rozloženie vzorov.

III. Operácie FEBFC Fuzzy Klasifikátora