



**WYŻSZA SZKOŁA
INFORMATYKI i ZARZĄDZANIA**
z siedzibą w Rzeszowie

Wyższa Szkoła Informatyki i Zarządzania
Kolegium Informatyki Stosowanej, kierunek Programowanie

Mariia Volokhonska w67971

Laboratorium 2 - Algorytm LEM + ocena regul, walidacja

Rzeszow 2024

1 Zadanie 1 (LEM) - Opis problemu

W tej pracy laboratoryjnej rozpatrzono problem związany z dokładnym określeniem stopnia zużycia paliwa w zależności od określonych charakterystyk samochodu, takich jak typ samochodu, jego cena oraz predkość. Do rozwiązania postawionego problemu należało skorzystać z algorytmu LEM, który doprowadzi do stworzenia reguł decyzyjnych, które pozwolą jednoznacznie określić stopień zużytego paliwa. Głównym zadaniem technicznym tego laboratorium było ustalenie powyższych reguł logicznych.

2 Badane zbiory danych

Tabela przedstawiona w tym laboratorium składa się ze zbioru przypadków dotyczących samochodów i decyzji o ustaleniu stopnia zużycia paliwa oraz jest zbiorem uczącym, po analizie którego możliwe będzie zbudowanie reguł decyzyjnych. Atrybutami tej tabeli są Typ, Cena, Predkość oraz Zużycie paliwa.

Typ: opisuje kategorie rozmiaru samochodu - "duży", "kompakt", "mały".

Cena: określa kategorie cenowa samochodu - "akceptowalna", "wysoka".

Predkość: określa kategorie predkości, do której można zaliczyć samochód - "przeciętna", "duża" oraz "wolna".

Zużycie paliwa: wskazuje na ilość paliwa zużytego przez samochód, przypisując samochodowi odpowiednią kategorię zużycia paliwa: średnie, wysokie, małe.

Typ	Cena	Predkość	Zużycie paliwa	
duży	akceptowalna	przeciętna	średnie	1
duży	wysoka	przeciętna	wysokie	2
duży	wysoka	duża	wysokie	3
kompakt	akceptowalna	przeciętna	średnie	4
kompakt	wysoka	duża	średnie	5
kompakt	wysoka	wolna	średnie	6
mały	akceptowalna	przeciętna	małe	7
mały	wysoka	mała	średnie	8
mały	akceptowalna	mała	małe	9
duży	wysoka	mała	wysokie	10

Rys. 1: Tabela z danymi - LEM

3 Obliczenia(dla Reguły 1)

1. Wypisano zbiory przypadków dla wszystkich klas decyzyjnych. B(klasa decyzyjna)
2. Wypisano zbiory wszystkich par ([Atrybut, Wartosc]). T(G)
3. Zbioru G przypisano wartości zbioru przypadków klasy decyzyjnej "średnie". $G=B(\text{średnie})=1,4,5,6,8$
4. Dla każdego zbioru par ([Atrybut, Wartosc]) wyznaczona ilość przypadków ze zbiorów, zawierających w zbiorze G oraz ilość pozostałych przypadków w tych zbiorach, które nie znajdują się w zbiorze G.
5. Znałeziono pare ([Atrybut, Wartosc]) z maksymalną ilością odpowiadających przypadków i minimalną nie odpowiadających. Ta pare oznaczono jako W1. [(Typ, kompakt)]=4,5,6 - 3 i 0.
6. Zbiór przypadków pary [(Typ, kompakt)]=4,5,6 należy do zbioru $G=1,4,5,6,8$, więc została utworzona reguła 1.
7. Od zbioru G odjęto pokryte przez Reguła1 przypadki dla kontynuacji obliczeń.
 $G=B(\text{średnie})-4,5,6=1,4,5,6,8-4,5,6=1,8$.

B(średnie)=[1,4,5,6,8]			
B(małe)=[7,9]			
B(wysokie)=[2,3,10]			
T(G) zbiór wszystkich par atrybut wartość			
[(Typ, duży)]=[1,2,3,10]			
[(Typ, kompakt)]=[4,5,6]			
[(Typ, mały)]=[7,8,9]			
[(Cena, akceptowalna)]=[1,4,7,9]			
[(Cena, wysoka)]=[2,3,5,6,8,10]			
[(Prędkość, duża)]=[3,5]			
[(Prędkość, mała)]=[8,9,10]			
[(Prędkość, przeciętna)]=[1,2,4,7]			
[(Prędkość, wolny)]=[6]			
G=B(średnie)=[1,4,5,6,8]			
T(G) zbiór wszystkich par atrybut wartość			
	[(Typ, duży)]=[1,2,3,10]	1	3
W1	[(Typ, kompakt)]=[4,5,6]	3	0
	[(Typ, mały)]=[7,8,9]	1	2
	[(Cena, akceptowalna)]=[1,4,7,9]	2	2
	[(Cena, wysoka)]=[2,3,5,6,8,10]	3	3
	[(Prędkość, duża)]=[3,5]	1	1
	[(Prędkość, mała)]=[8,9,10]	1	2
	[(Prędkość, przeciętna)]=[1,2,4,7]	2	2
	[(Prędkość, wolny)]=[6]	1	0
W1=[4,5,6] należy do G=[1,4,5,6,8], więc można stworzyć regułę			
Reguła 1: JEŻELI Typ=kompakt TO Zużycie paliwa=średnie (4,5,6)			
G=B(średnie)-[4,5,6]=[1,4,5,6,8]-[4,5,6]=[1,8]			

Rys. 2: Obliczenia Reguły1

Reguła 1: JEŻELI Typ=kompakt TO Zużycie paliwa=średnie
 Reguła 1 pokrywa takie przypadki: 4,5,6

4 Wyniki i analiza

Ilość powstałych reguł: **5**

Ilość przypadków: **10**

4.1 Zestaw otrzymanych w wyniku spełnienia algorytmu LEM reguł decyzyjnych

Reguła 1: JEŻELI Typ=kompakt TO Zużycie paliwa=średnie

Reguła 2: JEŻELI Typ=mały ORAZ Predkość=mała ORAZ Cena=wysoka TO Zużycie paliwa=średnie

Reguła 3: JEŻELI Typ=duży ORAZ Cena=akceptowalna TO Zużycie paliwa=średnie

Reguła 4: JEŻELI Typ= duży ORAZ Cena=wysoka TO Zużycie paliwa=wysokie

Reguła 5: JEŻELI Typ=mały ORAZ Cena=akceptowalna TO Zużycie paliwa=małe

4.2 Ilość przypadków, pokrytych przez reguły

Reguła 1: 3

Reguła 2: 1

Reguła 3: 1

Reguła 4: 3

Reguła 5: 2

4.3 Reguły, pokrywające maksymalna i minimalna ilość przypadków

Maksymalna ilość: Reguła 1 oraz Reguła 4 (3 przypadki)

Minimalna ilość: Reguła 2 oraz Reguła 3 (1 przypadek)

4.4 Reguły, pokrywające maksymalna i minimalna ilość przypadków

Maksymalna ilość: Reguła 1 oraz Reguła 4 (3 przypadki)

Minimalna ilość: Reguła 2 oraz Reguła 3 (1 przypadek)

4.5 Przypadki klasyfikowane przez reguły

Reguła 1: 4, 5, 6

Reguła 2: 8

Reguła 3: 1

Reguła 4: 2, 3, 10

Reguła 5: 7, 9

4.6 Specyfika reguł decyzyjnych

Najbardziej ogólna reguła: Reguła 1, zawiera tylko 1 warunek.

Najbardziej szczegółowa reguła: Reguła 2, zawiera 3 warunki.

4.7 Atrybuty jako warunki

Najbardziej często jako jeden z warunków reguły występował atrybut Typ.(jest częścią wszystkich reguł). Najmniej często występowała Predkość(jest częścią reguły 2).

5 Wnioski

W wyniku przeprowadzonej pracy laboratoryjnej dotyczącej wykorzystania algorytmu LEM dla ustalenia stopnia zużycia paliwa przez samochód w oparciu o jego cechy, można wyciągnąć następujące wnioski:

1. Mówiąc o samym zbiorze danych, można powiedzieć, że najwięcej przypadków należy do klasy decyzji "średnie", dlatego większość reguł (3 z 5) ma na celu ustalenie, czy zużycie paliwa jest średnie.
2. Jeżeli wyciągnąć wnioski na temat reguł decyzyjnych uzyskanych podczas pracy laboratoryjnej, to można powiedzieć, że reguły pomyślnie pokryły wszystkie przypadki danego zbioru. Oznacza to, że zgodnie z ustalonymi regułami nie będzie w przyszłości problemów z klasyfikacją nieprzewidzianych przypadków.
3. Jako rezultat danej pracy laboratoryjnej zostały stworzone 5 reguł decyzyjnych. Pierwsza mówi o tym, że zawsze jeżeli typ samochodu kompakt, to zużycie paliwa jest średnie, dlatego jeżeli człowiek wie z góry, że jest gotowy wydać umiarkowane kwoty na utrzymanie samochodu, wtedy lepiej dla niego od razu pomyśleć o samochodzie kompaktowym. Reguła 2 mówi o tym, że jeżeli typ samochodu mały oraz ma małą predkość oraz cena na niego jest wysoka, to zużycie paliwa takiego samochodu jest średnie. Ta informacja daje możliwość człowiekowi, kupującemu samochód pomyśleć nad tym, czy jest sens taki samochód kupować, ponieważ i charakterystyki samochodu są nie bardzo przyjemne, i zużycie paliwa przy tym będzie średnie, a nawet nie małe. Reguła 3 mówi, że jeżeli typ jest duży oraz cena jest akceptowalna to zużycie paliwa jest średnie. Reguła 4 pomaga jednoznacznie ustalić, że duży samochód za wysoką cenę ma wysokie zużycie paliwa. To oznacza, że osoba, która kupiła mocny, drogi samochód, musi być przygotowana na zapłacenie dużych pieniędzy za jego utrzymanie. Reguła 5 mówi o tym, że jeżeli typ samochodu jest mały, a cena akceptowalna, to zużycie paliwa będzie małym, czyli człowiek, który nie posiada dużo pieniędzy utrzymanie samochodu, może skorzystać z tej reguły i zatrzymać swój wybór na małym tanim samochodzie.
4. Jeśli mówić o algorytmie LEM, to można powiedzieć, że algorytm LEM jest doskonałym narzędziem do tworzenia reguł decyzyjnych, a w porównaniu z algorytmem GTS jest znacznie wygodniejszy w użyciu z punktu widzenia braku bezpośrednich obliczeń arytmetycznych i z punktu widzenia koncentracji na pracy ze zbiorami.

Podsumowując powyższe można stwierdzić, że stworzone w procesie wykonywania danego laboratorium reguły ułatwia życie nabywcom samochodów i pozwala im szybko przeanalizować swoje możliwości, a najlepszy sposób otrzymania tych reguł jest korzystanie z algorytmu LEM, ponieważ jest on nie taki czasochłonny, jak GTS oraz łatwiejszy w użyciu.

6 Zadanie 2 (ocena reguł, walidacja) - Opis problemu

W tym laboratorium rozpatrzono problem związany z potrzebą oceny efektywności, dokładności i poprawności reguł decyzyjnych podanych w tym laboratorium, w oparciu o określony zestaw danych. Podczas pracy laboratoryjnej każda reguła będzie oceniana według następujących parametrów oceny reguły: Siła, Dokładność, Ogólność, Specyficzność, Wsparcie.

7 Badane zbiory danych

7.1 Zestaw reguł

REGULA 1

JEZELI Lzawienie JEST zmniejszone
TO SOCZEWKI JEST brak

REGULA 2

JEZELI Wiek JEST młody
ORAZ Astygmatyzm JEST nie
ORAZ Lzawienie JEST normalne
TO SOCZEWKI JEST miękkie

REGULA 3

JEZELI Wada wzroku JEST krótkowidz
ORAZ Astygmatyzm JEST tak
ORAZ Lzawienie JEST normalne
TO SOCZEWKI JEST twarde

REGULA 4

JEZELI Wiek JEST podeszały
ORAZ Astygmatyzm JEST nie
ORAZ Lzawienie JEST normalne
TO SOCZEWKI JEST miękkie

REGULA 5

JEZELI Wiek JEST młody
ORAZ Astygmatyzm JEST tak
ORAZ Lzawienie JEST normalne
TO SOCZEWKI JEST twarde

REGULA 6

JEZELI Wiek JEST młody
ORAZ Astygmatyzm JEST tak
ORAZ Lzawienie JEST normalne
TO SOCZEWKI JEST twarde

7.2 Tabela z danymi - ocena reguł

ZBIÓR DANYCH					
LP	Wiek	Wada	Astyg	Lzawienie	Soczew
1	młody	dalekowidz	nie	zmniejszone	brak
2	młody	krotkowidz	tak	zmniejszone	brak
3	prestarczy	krotkowidz	tak	zmniejszone	brak
4	prestarczy	krotkowidz	nie	normalne	miekkie
5	młody	krotkowidz	tak	normalne	twarde
6	starczy	dalekowidz	tak	zmniejszone	brak
7	prestarczy	dalekowidz	nie	zmniejszone	brak
8	prestarczy	dalekowidz	tak	zmniejszone	brak
9	prestarczy	krotkowidz	nie	zmniejszone	brak
10	młody	dalekowidz	tak	zmniejszone	brak
11	starczy	krotkowidz	tak	normalne	twarde
12	prestarczy	dalekowidz	nie	normalne	miekkie
13	starczy	dalekowidz	tak	normalne	brak
14	starczy	krotkowidz	tak	zmniejszone	brak
15	starczy	krotkowidz	nie	normalne	brak
16	prestarczy	krotkowidz	tak	normalne	twarde
17	młody	krotkowidz	nie	zmniejszone	brak
18	starczy	krotkowidz	nie	zmniejszone	brak
19	starczy	dalekowidz	nie	zmniejszone	brak
20	młody	dalekowidz	nie	normalne	miekkie
21	starczy	dalekowidz	nie	normalne	miekkie
22	prestarczy	dalekowidz	tak	normalne	brak

Rys. 3: Tabela z danymi - ocena

8 Obliczenia

1. **Sila** = E_c , gdzie E_c to liczba poprawnie klasyfikowanych przypadków ze zbioru uczącego..
2. **Dokładność** = $E_c / (E_c + E_e)$, gdzie $E_c + E_e$ to suma poprawnie i nie poprawnie klasyfikowanych przypadków ze zbioru uczącego.(z jakim prawdopodobieństwem przypadek w zbiorze uczącym może okazać się poprawnie klasyfikowanym?)
3. **Ogólność** = $(E_c + E_e) / E$, gdzie E - to liczba wszystkich przypadków w zbiorze.(w jakim stopniu przypadki z wybranego zbioru warunków, ponieważ zbiór warunków daje poprawnie i nie poprawnie klasyfikowane przypadki, występują w ogólnym zbiorze).
4. **Specyficzność** = E_c / E_{class} , gdzie E_{class} - to liczba wszystkich przypadków w danej klasie decyzyjnej.(stosunek liczby poprawnie klasyfikowanych przypadków do liczby przypadków z wybranej klasy decyzyjnej)
5. **Wsparcie** = E_c / E , gdzie E_{class} - to liczba wszystkich przypadków w danej klasie decyzyjnej.(stosunek poprawnie klasyfikowanych przypadków do wszystkich przypadków zbioru).