**WYŻSZA SZKOŁA INFORMATYKI**

**I ZARZĄDZANIA**

Z SIEDZIBĄ W RZESZOWIE

Sprawozdanie

Sztuczna inteligencja

**Optymalizacja modeli uczenia**

Prowadzący: dr inż. Mariusz Wrzesień Wykonawca: Sylwia Babiarz 46741

6IID – GAK, SL04

Rzeszów 2014

# Opis problemu

Celem laboratorium jest optymalizacja modeli uczenia, przy pomocy danych umiejscowionych w tablicy decyzji. Po wygenerowaniu reguł należy ocenić ich zbiór, gdyż może się okazać, że niektóre z nich są zbędne bądź można je w pewien sposób poprawić.

# Badane zbiory danych

Dane, jakie należy przetworzyć, dotyczą problemu doboru soczewek. Są one przedstawione za pomocą tablicy decyzji, której skład to pięć kolumn i dwadzieścia dwa wiersze. Wszystkie wiersze to przypadki, jakie są rozpatrywane. Kolumna pierwsza jest indeksowana atrybutem porządkowym o nazwie „wiek”. Zawiera trzy wartości atrybutów, tj. młody, starczy i prestarczy. Kolumna druga indeksowana jest atrybutem nominalnym „wada wzroku”. Posiada dwie wartości, dalekowidz oraz krótkowidz. Kolejna kolumna ma nazwę atrybutu nominalnego „astygmatyzm” i dysponuje także dwoma wartościami, tak i nie. Przedostatnia z kolumn indeksowana jest atrybutem porządkowym „łzawienie” wraz z dwoma wartościami, normalne i zmniejszone. Ostatnia kolumna ma nazwę atrybutu „soczewki” i jest to kolumna decyzji. Składa się z trzech klas: brak, miękkie, twarde. Kolumna Lp. nie zawiera się w ciele tablicy decyzji.

Tabela 1 Tablica decyzji problemu doboru soczewek

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| LP. | **WIEK** | **WADA WZROKU** | **ASTYGMATYZM** | **LZAWIENIE** | **SOCZEWKI** |
| 1 | młody | dalekowidz | nie | zmniejszone | brak |
| 2 | młody | krótkowidz | tak | zmniejszone | brak |
| 3 | prestarczy | krótkowidz | tak | zmniejszone | brak |
| 4 | prestarczy | krótkowidz | nie | normalne | miękkie |
| 5 | młody | krótkowidz | tak | normalne | twarde |
| 6 | starczy | dalekowidz | tak | zmniejszone | brak |
| 7 | prestarczy | dalekowidz | nie | zmniejszone | brak |
| 8 | prestarczy | dalekowidz | tak | zmniejszone | brak |
| 9 | prestarczy | krótkowidz | nie | zmniejszone | brak |
| 10 | młody | dalekowidz | tak | zmniejszone | brak |
| 11 | starczy | krótkowidz | tak | normalne | twarde |
| 12 | prestarczy | dalekowidz | nie | normalne | miękkie |
| 13 | starczy | dalekowidz | tak | normalne | brak |
| 14 | starczy | krótkowidz | tak | zmniejszone | brak |
| 15 | starczy | krótkowidz | nie | normalne | brak |
| 16 | prestarczy | krótkowidz | tak | normalne | twarde |
| 17 | młody | krótkowidz | nie | zmniejszone | brak |
| 18 | starczy | krótkowidz | nie | zmniejszone | brak |
| 19 | starczy | dalekowidz | nie | zmniejszone | brak |
| 20 | młody | dalekowidz | nie | normalne | miękkie |
| 21 | starczy | dalekowidz | nie | normalne | miękkie |
| 22 | prestarczy | dalekowidz | tak | normalne | brak |

# Obliczenia

Do powyższej tablicy decyzji zostało już wcześniej wygenerowane sześć reguł. Brzmią one następująco:

* R1 JEZELI Łzawienie JEST zmniejszone TO SOCZEWKI JEST brak
* R2 JEZELI Wiek JEST młody ORAZ Astygmatyzm JEST nie ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST miękkie
* R3 JEZELI Wada\_wzroku JEST krótkowidz ORAZ Astygmatyzm JEST tak ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST twarde
* R4 JEZELI Wiek JEST prestarczy ORAZ Astygmatyzm JEST nie ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST miękkie
* R5 JEZELI Wiek JEST młody ORAZ Astygmatyzm JEST tak ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST twarde
* R6 JEZELI Wiek JEST młody ORAZ Astygmatyzm JEST tak ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST twarde .

Kolejnym krokiem jest wykonanie dla tych reguł obliczeń ustalających jaką siłę, dokładność, ogólność, specyficzność i wsparcie ma każda z nich.

Parametr siła określa liczbę poprawnie klasyfikowanych przypadków. Siła = Ep.

Parametr dokładność jest stosunkiem liczby poprawnie klasyfikowanych przypadków (Ep) z bazy informacyjnej do sumy przypadków klasyfikowanych poprawnie (Ep) oraz błędnie (Eb).

Dokładność =

Parametr ogólność jest stosunkiem sumy przypadków klasyfikowanych poprawnie (Ep) oraz błędnie (Eb) do całkowitej liczby przypadków w zbiorze uczącym (E).

Ogólność =

Parametr specyficzność jest stosunkiem liczby poprawnie klasyfikowanych przypadków (Ep) z bazy informacyjnej do liczby przypadków w tej bazie, należących do danej klasy decyzyjnej (Ek).

Specyficzność =

Parametr wsparcie to stosunek liczby poprawnie klasyfikowanych przypadków (Ep) z bazy informacyjnej do całkowitej liczby przypadków w tym zbiorze (E).

Wsparcie =

Dla powyżej przedstawionych reguł przygotowano tabelę wraz z obliczonymi parametrami i pokrytymi przez nie przypadkami.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Reguła | Siła | Dokładność | Ogólność | Specyficzność | Wsparcie | Przypadki |
| R1 | 12 | 1 | 0,545455 | 0,8 | 0,545455 | 1, 2, 3, 6, 7, 8, 9, 10, 14, 17, 18, 19 |
| R2 | 1 | 1 | 0,045455 | 0,25 | 0,045455 | 20 |
| R3 | 3 | 1 | 0,136364 | 1 | 0,136364 | 5, 11, 16 |
| R4 | 1 | 1 | 0,090909 | 0,5 | 0,090909 | 4, 12 |
| R5 | 1 | 1 | 0,045455 | 0,33 | 0,045455 | 5 |
| R6 | 1 | 1 | 0,045455 | 0,33 | 0,045455 | 5 |

Jak można zauważyć, nie wszystkie przypadki tablicy decyzji zostały pokryte przez sześć powyższych reguł. W związku z tym, w kolejnym kroku, należy wyznaczyć kolejne reguły przy pomocy algorytmu GTS.

Aby wygenerować pierwszą regułę, w pierwszym kroku algorytmu zostaje wzięty pod uwagę pierwszy, niepokryty regułą przypadek, czyli wiersz o liczbie porządkowej 13. Wartość dla pierwszego atrybutu to *Wiek = starczy* z odpowiadającą mu w kolumnie decyzji klasą *Soczewki= brak*. Za pomocą filtrów w programie Microsoft Excel szukane są inne przypadki z taką samą kombinacją. Kolejnym krokiem jest obliczenie wartości parametrów, potrzebnych do stworzenia reguł, opisanych wzorami:

* ogólność: ,
* dokładność ,
* parametr H ,

gdzie

* Ep to liczba przypadków pozytywnych, czyli takich, które posiadają taką samą wartość atrybutu i decyzji w danym zestawieniu,
* Eb to liczba przypadków błędnych, czyli takich, które dla danej wartości atrybutu posiadają inną wartość decyzji,
* E to liczba wszystkich przypadków w tablicy decyzji.

Aby utworzyć regułę wartość parametru *A* musi być równa 1.

Bazując na powyższym opisie wszystkich przypadków (*E*) jest dwadzieścia dwa, dla pary *Wiek = starczy* oraz *Soczewki = brak* można odszukać sześć przypadków pozytywnych (*Ep*) oraz dwa przypadki błędne (*Eb*). Podstawiając dane do wzoru dla kolejnych wartości atrybutów w pierwszym przypadku, otrzymano:

Dla *Wiek =starczy*

Ponieważ parametr *A* nie jest równy 1, algorytm nadal przeszukuje tablicę decyzji biorąc pod uwagę kolejną cechę, tak więc:

Dla *Wada wzroku = dalekowidz*

Dla *Astygmatyzm = tak*

Dla *Łzawienie = normalne*

Ponieważ dla żadnej wartości atrybutu wartości decyzji *Soczewki = brak* wartość parametru dokładności *A=1*, nie można utworzyć reguły. Dlatego do cechy z najwyższym parametrem H dobierana jest kolejna i analizowana jest kolejna kombinacja. Dla tego przypadku cechy, jakie należy połączyć to: Wada wzroku = dalekowidz oraz Astygmatyzm = tak, reguła ta brzmi następująco:

JEŻELI Wada wzroku = dalekowidz ORAZ Astygmatyzm = tak TP Soczewki = brak

Reguła ta ma zastosowanie w przypadku 13 i 22, a także 6, 8, i 10, które były już pokryte regułą nr 1.

Dla tej reguły, także konieczne jest obliczenie siły, dokładności, ogólności, specyficzności i wsparcia. W tym przypadku konieczne jest wygenerowanie kolejnych reguł tak, aby cały zbiór był nimi pokryty.

# Wyniki

* R1 JEZELI Łzawienie JEST zmniejszone TO SOCZEWKI JEST brak
* R2 JEZELI Wiek JEST młody ORAZ Astygmatyzm JEST nie ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST miękkie
* R3 JEZELI Wada\_wzroku JEST krótkowidz ORAZ Astygmatyzm JEST tak ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST twarde
* R4 JEZELI Wiek JEST prestarczy ORAZ Astygmatyzm JEST nie ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST miękkie
* R5 JEZELI Wiek JEST młody ORAZ Astygmatyzm JEST tak ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST twarde
* R6 JEZELI Wiek JEST młody ORAZ Astygmatyzm JEST tak ORAZ Łzawienie JEST normalne TO SOCZEWKI JEST twarde
* R7 JEŻELI Wada wzroku = dalekowidz ORAZ Astygmatyzm = tak TO Soczewki = brak
* R8 JEŻELI Wiek = starczy ORAZ Wada wzroku = krótkowidz ORAZ Astygmatyzm = nie TO Soczewki = brak
* R9 JEŻELI Wada Wzroku = dalekowidz ORAZ Astygmatyzm = nie ORAZ Łzawienie = normalne TO Soczewki = miękkie

# Wnioski

Celem laboratorium była optymalizacja modeli uczenia, a więc optymalizacja reguł. Pierwszym krokiem było degenerowanie brakujących reguł, przy pomocy algorytmu GTS. Ocena zbioru zależy od ich ilości, liczby warunków w regułach, dokładności zbioru i liczby błędnie sklasyfikowanych przypadków.

W przypadku powyższej tablicy, już na pierwszy rzut oka widać, że liczba reguł może zostać pomniejszona. Reguła numer 6 oraz reguła o numerze 5 jest złożona z tych samych kombinacji atrybutów, dlatego jedną z nich można usunąć. Jest to reguła redundantna. Ponadto, dla danych zawartych w tej tablicy, reguły 5 i 6 nie są w ogóle potrzebne, ponieważ jedyny przypadek jaki pokrywają jest pokryty przez wcześniejszą regułę o numerze 3. Podobnie rzecz się ma z regułą 2, jedyny przypadek jaki jest przez nią pokryty jest także zawarty w regule numer 9, która ma większą istotność. Tak więc zbiór może zostać zoptymalizowany z 9 reguł do 6 reguł.

Średnia liczba warunków w regułach wynosi w przybliżeniu 2,67. Większość reguł składa się z trzech warunków.

Dokładność zbioru reguł, czyli liczba poprawnie klasyfikowanych przypadków do liczby wszystkich klasyfikowanych przypadków w tym przykładzie zawsze jest równa 1.

Błąd klasyfikacji nie dotyczy powyższego przykładu, ponieważ wszystkie przypadki są klasyfikowane poprawnie. Jest to dosyć istotnym faktem przy systemach sztucznej inteligencji oraz tworzeniu reguł.