**WYŻSZA SZKOŁA INFORMATYKI**

**I ZARZĄDZANIA**

Z SIEDZIBĄ W RZESZOWIE

Sprawozdanie

Sztuczna inteligencja

**Regułowe systemy decyzyjne GTS**

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

6IID – GAK, SL04

Rzeszów 2014

# Opis problemu

Celem laboratorium jest utworzenie zestawu reguł decyzyjnych, dla danych umiejscowionych w tablicy decyzji, przy pomocy algorytmu GTS. Reguły składają się z części warunkowej oraz decyzyjnej. Ich formalna reprezentacja jest wykorzystywana zarówno w uczeniu maszynowym, jak i odkrywaniu wiedzy z danych, dlatego tak ważne jest by zrozumieć jak powstają.

# 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.

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

Na przykładzie bazy informacyjnej z problemem doboru soczewek można opisać działanie algorytmu GTS. Aby wygenerować pierwszą regułę w pierwszym kroku algorytmu zostaje wzięty pod uwagę pierwszy, niepokryty regułą przypadek, czyli wiersz o liczbie porządkowej 1. Wartość dla pierwszego atrybutu to *Wiek = młody* 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 = młody* oraz *Soczewki = brak* można odszukać cztery przypadki pozytywne (*Ep*) oraz sześć przypadków błędnych (*Eb*). Podstawiając dane do wzoru dla kolejnych wartości atrybutów w pierwszym przypadku, otrzymano:

Dla *Wiek = młody*

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

Dla *Wada wzroku = dalekowidz*

Dla *Astygmatyzm = nie*

Dla *Łzawienie = zmniejszone*

Ponieważ dla wartości atrybutu *Łzawienie = zmniejszone* i wartości decyzji *Soczewki = brak* wartość parametru dokładności *A=1*, można utworzyć regułę 1, która brzmi:

**JEŻELI** Łzawienie = zmniejszone **TO** Soczewki = brak

Powyższa reguła pokrywa przypadki {1, 2, 3, 6, 7, 8, 9, 10, 14, 17, 18, 19}. Zatem kolejnym rozpatrywanym przypadkiem będzie następny nie pokryty przez żadną regułę przypadek, czyli przypadek 4, który należy badać w ten sam sposób. Jeżeli w którymś z badanych przypadków występuje sytuacja, gdzie żadna z kombinacji wartości pojedynczych atrybutów z daną klasą nie osiąga wartości *A=1* to należy wybrać tą, która miała najwyższy parametr *H* i dobierać do niej kolejne kombinacje, aż parametr *A* będzie miał odpowiednią wartość.

# Wyniki

W wyniku opisanego powyżej algorytmu powstało sześć reguł pokrywające wszystkie przypadki:

1. JEŻELI łzawienie = zmniejszone TO soczewki = brak

Dla przypadków: {1, 2, 3, 6, 7, 8, 9, 10, 14, 17, 18, 19}

1. JEŻELI wiek = prestarczy ORAZ astygmatyzm = nie ORAZ łzawienie = normalne TO soczewki = miękkie

Dla przypadków: {4, 12}

1. JEŻELI wada wzroku = krótkowidz ORAZ astygmatyzm = tak ORAZ łzawienie = normalne TO soczewki = twarde

Dla przypadków: {5, 11, 16}

1. JEŻELI wada wzroku = dalekowidz ORAZ astygmatyzm = tak TO soczewki = brak

Dla przypadków: {13, 22}

1. JEŻELI wiek = starczy ORAZ wada wzroku = krótkowidz ORAZ astygmatyzm = nie TO soczewki = brak

Dla przypadka: {15}

1. JEŻELI wada wzroku = dalekowidz ORAZ astygmatyzm = nie ORAZ łzawienie = normalne TO soczewki = miękkie

Dla przypadków: {20, 21}

# Wnioski

Algorytm GTS pracuje na zasadzie sekwencyjnego pokrywania, czyli gdy wygeneruje pierwszą regułę i znajdzie dla niej przypadki, to w następnym cyklu pomija wszystkie przypadki pozytywne dla tej reguły. Następnie algorytm analizuje pozostałe przypadki zamieszczone w niesprzecznej bazie informacyjnej. Wynik to zbiór reguł, które pokrywają wszystkie przypadki ze zbioru uczącego. Algorytm nie zakończy swojej pracy, dopóki pewien przypadek nie jest pokryty regułą.

Z powyższej bazy wiedzy powstało sześć reguł. Pierwsza reguła pokrywa aż dwanaście przypadków, natomiast reguła numer pięć pokrywa tylko jeden przypadek. Ilość pokrytych przypadków zależy więc od podobieństwa, częstotliwości innych obiektów. Niektóre przypadki mogą być opisane przez kilka reguł. Na przykład przypadek numer 12 jest opisany regułą 2, ale reguła numer 6 także się dla niego sprawdza. Ułożenie przypadków ma znaczenie dla kolejności generowania reguł.

Uważam, że badana baza wiedzy jest zbyt mała aby określić niezawodność i optymalność algorytmu GTS. Jednak przy bardziej złożonej sytuacji, gdy atrybutów jest wiele, może on być skomplikowany do obliczenia reguł.