

Marcin Surkont
nr 87167
msurkont@box43.pl

Wrocław, 26 maja 2001

Automatyczne pozyskiwanie wiedzy – dr Kwaśnicka

Program dydaktyczny – algorytm AQ

Abstrakt

Celem projektu było stworzenie programu dydaktycznego objaśniającego działanie algorytmu AQ, będącego przedstawicielem algorytmów indukcyjnego uczenia się pojęć.

W ramach projektu przygotowany został program wykonujący algorytm AQ na podanym zbiorze danych oraz prezentacja multimedialna, prezentująca sam algorytm.

Program umożliwia wykonanie algorytmu krok po kroku dla danych wczytanych z pliku oraz sprawdzenie własnego przykładu na wynikowym zbiorze reguł.

Prezentacja multimedialna umożliwia zapoznanie się trzema głównymi częściami tego algorytmu, wykorzystywanymi definicjami oraz sprawdzenie wiadomości użytkownika przez wypełnienie krótkiego, 3-pytaniowego testu.

Data oddania – 28 maja 2001

Cel projektu

Celem projektu jest przygotowanie programu dydaktycznego, który prezentowałby działanie algorytmu AQ.

Wprowadzenie teoretyczne

Algorytm AQ należy do algorytmów indukcyjnego uczenia się pojęć. Wykorzystuje on zbiór reguł do prezentowania określonej hipotezy dla zdefiniowanego zestawu atrybutów. Jest przedstawicielem algorytmów sekwencyjnego pokrywania zbioru uczącego.

Opis algorytmu znajduje się w prezentacji multimedialnej.

Opis projektu

Z uwagi na dość duży stopień skomplikowania algorytmu oraz, z tego powodu, trudność w prześledzeniu jego działania nawet na najmniejszym zbiorze uczącym, przyjęte zostało następujące rozwiązanie: przygotowany został program, który pozwala na uzyskanie wynikowego zbioru reguł z algorytmu AQ, z minimalną informacją na temat działania algorytmu, oraz prezentacji multimedialnej, umożliwiającej zapoznanie się z podstawowymi częściami algorytmu.

Prezentacja multimedialna została rozszerzona o możliwość przetestowania znajomości algorytmu poprzez wypełnienie krótkiego testu.

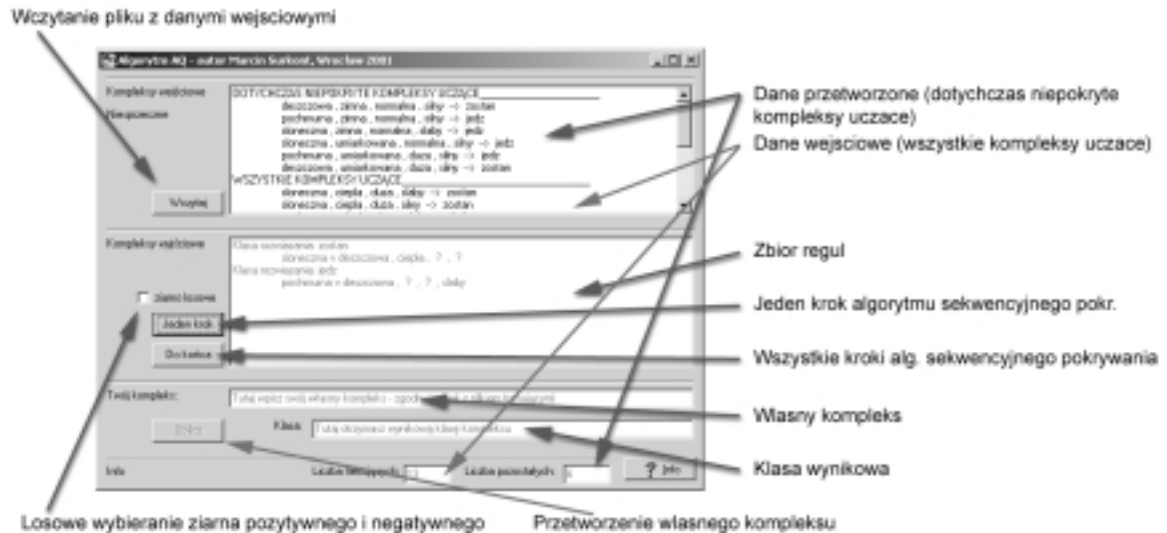
Program AQ

Interfejs

Początkowym etapem uruchomienia programu jest wybranie opcji „wczytaj”, co pozwoli na wybór pliku z danymi trenującymi. Następnie można wykonać operacje uruchamiające algorytm AQ: „Jeden krok” oraz „Do końca”.

Po skończeniu działania algorytmu możliwe jest sprawdzenie własnego kompleksu na wynikowym zbiorze reguł.

Interfejs programu przedstawia się następująco.



Dane wejściowe

Dane, które przetwarza program, muszą być przygotowane w odpowiednim formacie. Umieszczone powinny być w pliku z rozszerzeniem „.txt”.

Perwszą linią jest określenie liczby atrybutów i liczby wyjść (klas) – dwoma cyframi typu int (np.:”4 1”).

Każda następną linią jest oddzielnym kompleksem. Atrybuty są zapisywane tekstowo, z tym, że każdy z nich musi być jedną całością – znak typu „whitespace” jest traktowany jako rozdzielający dwa atrybuty. Należy również specjalnie zwrócić uwagę na spójność w nazewnictwie wartości atrybutów, gdyż w przypadku różnicy na jednym bajcie program traktuje to jako dwie różne wartości.

Ważne, aby na końcu pliku nie znalazła się pusta linia, gdyż spowoduje to błąd w programie.

Przykładowy plik:

```
3 1
ciepło silny słońce isc
zimno słaby słońce nie_isc
zimno mocny chmury nie_isc
```

Razem z programem dostarczany jest plik „pogoda.txt”, będący reprezentacją zbioru uczącego z książki P. Cichosza.

Opis działania

Program wczytuje plik wejściowy i próbuje na jego podstawie stworzyć zbiór kompleksów trenujących. W przypadku powodzenia operacji tworzenia zbioru zostaje wyświetlony komunikat „niesprzeczne”. Gdy jednak w zbiorze trenującym znajdują się sprzeczne kompleksy (czyli dla takich samych atrybutów różne klasy), program usuwa je ze zbioru trenującego oraz wyświetla komunikat z liczbą sprzecznych kompleksów.

Następnie, po zainicjowaniu akcji przez użytkownika, wykonuje algorytm sekwencyjnego pokrywania zbiorów dopóki nie zostanie pokryty każdy element ze zbioru. Wynikowy zbiór reguł, po każdej iteracji, jest wyświetlany na ekranie.

Prezentacja multimedialna

Prezentacja „Algorytm AQ” ma na celu zaprezentowanie działania algorytmu sekwencyjnego pokrywania AQ.

Użytkownik ma możliwość przechodzenia po kolejnych poziomach algorytmu, który jest mu prezentowany z wykorzystaniem grafiki, animacji oraz tekstu.

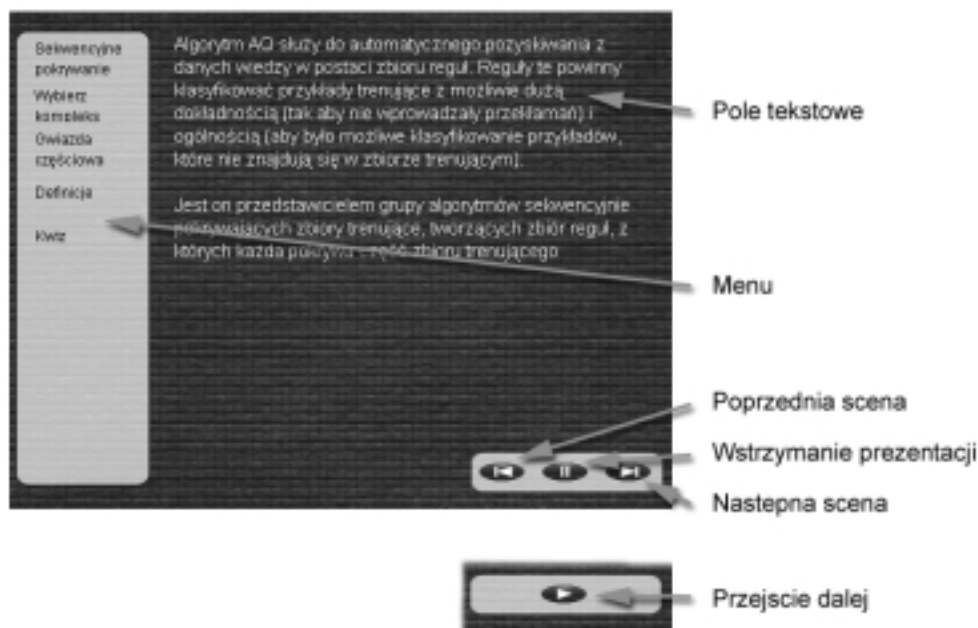
Miara sukcesu prezentacji

Po jednokrotnym oglądnięciu prezentacji użytkownik powinien być w stanie opisać zasadę działania algorytmu AQ.

Prezentacja

Prezentacja składa się ze wstępu, opisu algorytmu, definicji oraz kwizu.

Wstęp prezentuje nazwę algorytmu oraz nazwisko autora.



Prezentacja algorytmu AQ została podzielona na trzy poziomy (zgodnie z działaniem algorytmu):

- algorytm sekwencyjnego pokrywania
- funkcja wybrania kompleksu
- funkcja gwiazdy częściowej

Każdemu poziomowi opisu odpowiada zestaw scen, które go objaśniają. Każdy zestaw składa się ze strony tytułowej, opisującej krótko cel poziomu oraz jego umiejscowienie, oraz kilku stron opisujących poziom.

Definicje prezentują elementy opisane w prezentacji.

Narzędzie

Prezentacja multimedialna została przygotowana w systemie Macromedia Flash 5, co umożliwia umieszczenie jej na stronie www oraz przenoszenie jako niezależnego pliku na dyskietce.

Dyskusja autora

Działanie algorytmu AQ

Implementacja algorytmu AQ w programie umożliwiła sprawdzenie w praktyce zależności pomiędzy organizacją zbioru trenującego a wynikami działania algorytmu.

Rozmiar danych

Dla danych trenujących obejmujących dość duży obszar przestrzeni rozwiązań algorytm AQ zwraca reguły o dużej ogólności, obejmujące wszystkie przykłady trenujące.

W przypadku, gdy zbiór uczący obejmował jedynie niewielki wycinek przestrzeni rozwiązań, algorytm ten zwraca zaskakująco dużo reguł – często niewiele mniej, niż jest przykładów trenujących. Wiąże się to z istotą działania algorytmu – w przypadku stworzenia zbioru kompleksów pokrywających ziarno pozytywne bardzo prawdopodobne jest, że już przy pierwszych uszczegółowieniach tego zbioru nie będzie on pokrywał ziarna negatywnego, gdyż inne rozwiązania są bardzo od ziarna pozytywnego oddalone.

Kolejność danych

Ciekawy jest też olbrzymi rozrzut w liczbie generowanych reguł w wypadku włączenia losowego wybierania ziarna negatywnego i pozytywnego. W przypadku przykładowego pliku „pogoda.txt” liczba generowanych reguł wahała się od 4 do 7.

Różniła się też szczegółowość reguł – w wypadku małej ich liczby w regułach często występowały selektory uniwersalne <?>, natomiast w przypadku dużej liczby reguł większość z nich stanowiły kompleksy o atrybutach pojedynczych.

Podsumowanie

Cel projektu, stworzenie programu dydaktycznego prezentującego algorytm AQ został osiągnięty poprzez:

- stworzenie aplikacji pokazującej algorytm w działaniu, umożliwiającej analizę wynikowych zbiorów reguł
- stworzenie prezentacji multimedialnej, pozwalającej na zapoznanie się z działaniem algorytmu.

Dokładne zaprezentowanie działania algorytmu w prezentacji nie było możliwe, gdyż wymagałoby to w rzeczywistości przepisania opisu algorytmu z jednej z książek – a opisy te zajmują po kilka stron tekstu.

Dlatego też możliwość dokładnego zapoznania się z działaniem algorytmu została pozostawiona użytkownikowi.

Bibliografia

1. Paweł Cichosz „Systemy uczące się”, WNT Warszawa 2000
2. http://www.cee.hw.ac.uk/~howard/McL/McL_main.htm
3. <http://www.cs.bris.ac.uk/Teaching/Resources/COMS70300/slides/RulesHTML/sld001.htm>
4. <http://www.cse.unsw.edu.au/~cs9414/notes/ml/06prop/06prop.html>

Appendix 1

Plik scenariusza prezentacji multimedialnej.