

Salomon
System przetwarzania wiedzy

User's Manual, ver 1.1

Historia wersji

Data	Wersja	Opis	Autorzy
17-09-2005	1.0	User's Manual	Przemysław Misiuda, Krzysztof Nadolski, Ma- teusz Nowakowski, Łukasz Ostatek, Dominik Seweryn, Tomasz Wąsala
20-12-2005	1.0	Rozdział TreeConclude	Mateusz Nowakowski

1 Wstęp

Niniejszy dokument stanowi dokumentację użytkownika pomocną przy wykorzystaniu platformy Salomon do tworzenia drzew decyzyjnych.

1.1 Cel

Dokument ten powstał jako prosta instrukcja obsługi poszczególnych pluginów realizujących zadania kolejnych etapów procesu tworzenia i prezentowania drzew decyzyjnych. Wskazuje jakie opcje oferuje każdy plugin i jak ich używać oraz jakie należy wykonać kroki aby uzyskać pożądaną efekt w postaci drzewa decyzyjnego.

2 Pluginy

2.1 TreeDataLoader

Jest to pierwszy z pluginów tworzących drzewa decyzyjne. Służy do wybierania z bazy danych informacji potrzebnych do stworzenia drzewa. Użytkownik na początku wybiera tabelę na podstawie której będzie tworzone drzewo decyzyjne. Następnie kolumnę względem której tworzone będzie drzewo decyzyjne (decisioned column) oraz kolumny które mają być brane pod uwagę przy tworzeniu drzewa (decisioning columns).

Aby dokonać wyboru w/w tabeli i kolumn należy:

- załadować plugin
- dodać plugin do tasków
- kliknąć prawym przyciskiem na plugin (w oknie tasków) i wybrać „Settings”
- w oknie „Table” wybrać żadaną tabelę – wtedy w oknach „Decisioned column” oraz „Decioning columns” pojawi się lista dostępnych kolumn, zaś w polach "Range to" ilość wierszy w danej tabeli
- wybrać jedną kolumnę w „Decisioned kolumn” - lista kolumn w oknie „Decioning columns” powinna się zmiejszyć o wybraną kolumnę
- wybrać dowolną ilość w oknie „Decioning columns” (przy użyciu klawisza „Ctrl”)
- podać zakres wierszy danej tabeli (pola "Range from" i "Range to") które mają zostać użyte w algorytmie obliczeniowym
- kliknąć „OK”.

2.2 VeniTreeCreator

Jeśli chodzi o drugi w kolejności plugin - realizujący algorytm budowy drzew decyzyjnych - to uruchamia się go podobnie jak plugin pierwszy. Ze względu na swój charakter plugin ten nie podejmuje interakcji z użytkownikiem. Otrzymuje on od poprzedniego pluginu informacje na podstawie jakiej tabeli i w jaki sposób ma zbudować drzewo decyzyjne, a po skończeniu swej pracy przekazuje jej efekty do kolejnego tym razem wizualizacyjnego pluginu.

2.3 C45TreeCreator

Alternatywą dla opisanego powyżej pluginu algorytmicznego implementującego zasadę działania algorytmu ID3 jest plugin C45TreeCreator implementujący algorytm C4.5. W ustawieniach tego pluginu można podać poziom ufności na jakim będzie działał ten plugin. Wartość ta oznacza minimalny odsetek elementów znajdujących się w węźle drzewa aby węzeł ten został uznany za homogeniczny i nie rozwijany dalej. Sensowne wartości dla tego parametru są z przedziału $(0,5;1]$. Dodatkową cechą tego algorytmu jest, że potrafi on dla atrybutów ciągłych operować na ich zakresach zamiast na dyskretnych wartościach. Przykładowo zamiast tworzyć osobne węzły dla wartości atrybutu ciągłego wynoszących 1,2,3,4,5,6 może np. stworzyć dwa węzły dla wartości ≤ 2 i > 2 . U podstaw algorytm tworzy drzewa korzystając z obliczanego dla każdego atrybutu współczynnika GainRatio. Użycie tego współczynnika zamiast używanej w algorytmie ID3 entropii redukuje niekorzystny wpływ atrybutów posiadających wiele dopuszczalnych wartości. Dodatkowo GainRatio jest wyliczane tylko na danych, dla których wartość danego atrybutu jest określona. W ten sposób zredukowana jest ułomność algorytmu ID3, który to algorytm wymagał, aby wartości wszystkich atrybutów dla wszystkich elementów danych były określone.

2.4 TreeVisualisation

Jest to trzeci z grupy pluginów odpowiadających za tworzenie drzew decyzyjnych. Uruchamia się go analogicznie jak poprzednie pluginy. Służy do prezentacji graficznej utworzonych drzew. W pierwszej iteracji został on napisany jako bardzo prymitywny i nie ma możliwości wyboru ustawień pluginu. Aby obejrzeć drzewo należy uruchomić plugin, a następnie kliknąć na Result. Po naciśnięciu Result w okienku tekstowym pojawi się drzewo w postaci wypisanego stringu. Gdzie mamy najpierw korzeń drzewa, a później zestawienia rodzic, krawędź, potomek. Kolejne poziomy są oddzielone znakiem `"/`, natomiast poszczególne węzły na poziomie są oddzielone znakiem `"#"`.

Plugin może działać w dwóch trybach.

1. Jeśli chcemy aby działał razem z innymi pluginami i wyświetlił drzewo stworzone przez nie, nie trzeba nic zmieniać w jego ustawieniach, wystarczy odpowiednio

poustawiać wcześniejsze pluginy i kliknąć Run. W oknie Result pluginu pojawi się okno stworzone przez wcześniej wybrane pluginy.

2. Jeśli chcemy aby plugin działał samodzielnie, wchodzimy do panelu Settings. W panelu mamy możliwość skasowania już istniejącego drzewa. Wystarczy wybrać istniejące drzewo z listy i wcisnąć przycisk Usuń drzewo. Drzewo zostanie skasowane. Drugą funkcją, którą można wykorzystać w pluginie działającym samodzielnie jest rysowanie drzewa wybranego z listy, aby to zrobić należy wybrać interesujące nas drzewo z listy, zaznaczyć checkbox Działaj samodzielnie. Następnie należy kliknąć przycisk Run, wejść do okna Result gdzie pojawi się interesujące nas drzewo w postaci JTree. Jeśli zrobimy coś niezgodnie z instrukcją w oknie Result pojawi się komunikat z informacją jak należy postąpić z pluginem aby wyświetlić drzewo.

2.5 TreeConclude

Plugin służy do zweryfikowania drzewa tj. sprawdzenia czy na innych danych opisujących tą samą rzeczywistość drzewo zwraca prawidłowe rezultaty.

Plugin uruchamia się analogicznie jak pozostałe. Podobnie jak plugin wizualizacyjny może pracować w dwóch trybach: samodzielnie lub współdziałając z innymi pluginami. Tryb pracy określa się w oknie ustawień.

2.5.1 Okno ustawień

Okno ustawień definiuje drzewo, które będzie testowane, natomiast danymi testowymi są pozostałe wiersze w tabeli źródła danych, które nie uczestniczyły w definiowaniu drzewa np. jeżeli drzewo było zbudowane za pomocą wierszy 2-8, a tabela zawiera 14 wierszy wówczas test będzie przeprowadzony na wierszach: 1,9-14.

W oknie ustawień określa się za pomocą checkbox'a *Pracuj samodzielnie* tryb pracy. Domyślnym ustawieniem jest współdziałanie z poprzednimi pluginami tj. obecności w środkowisku Salomona wygenerowanego drzewa przez plugin algorytmiczny. W przypadku ustawienia pracy samodzielnej należy określić za pomocą listy wyboru *Wybierz drzewo* interesujące nas drzewo.

Po wybraniu konkretnego drzewa w oknie pokazują się niektóre właściwości drzewa tj. jego nazwa, opis, definicja źródła danych na podstawie którego drzewo powstało. Po wybraniu drzewa do testów, należy kliknąć *OK* w celu zatwierdzenia.

2.5.2 Okno rezultatów

Okno wyników przedstawia wynik testu tzn. nazwę drzewa, ilość wierszy na których przeprowadzono test, ilość testów pozytywnych, procent zgodności drzewa oraz listę wierszy, których wynik nie jest zgodny z definicją drzewa. Lista zawiera kolumny: nr

wiersza, wynik skonstruowany za pomocą drzewa, spodziewany wynik tj. taki jaki był w bazie danych, oraz kolumny decyzyjne.