

Systemy uczące się – Zaliczenie projektu

dr Przemysław Juszczuk

Katedra Inżynierii Wiedzy, Uniwersytet Ekonomiczny

I 2019

Zdefiniowanie problemu i wybór odpowiedniej metody

- Przegląd teoretyczny uwzględniający podstawowe informacje o algorytmie.
- Przegląd i krótki opis kluczowych elementów algorytmu mających wpływ na jego działanie (operatory genetyczne, przygotowanie wektora wejściowego, oczekiwane wyjście algorytmu).
- Uproszczony schemat algorytmu w postaci pseudokodu (opcjonalnie w postaci schematu graficznego).
- Analiza danych wejściowych. Sprawdzenie, czy konieczny jest preprocessing danych, usunięcie outlierów, uzupełnienie/usunięcie danych niepełnych, dyskretyzacja lub fuzyfikacja danych.
- Analiza prostego przypadku w postaci symulacji krokowej. Oczekiwany wynik powinien być znany.

Zdefiniowanie problemu i wybór odpowiedniej metody

- Analiza prostego przypadku w postaci symulacji krokowej. Oczekiwany wynik powinien być znany już na początku tak, aby prześledzić dokładnie, jak modyfikowane są wartości pośrednie w algorytmie. Takie działanie odpowiada pośrednio testom jednostkowym, gdzie każdą metodę możemy traktować jako black box.
- * Zapis formalny zadania, zapis formalny wybranych operatorów, mechanizmów zastosowanych w algorytmie.

Opracowanie części praktycznej

- * Przygotowanie klas abstrakcyjnych/interfejsów dla ogólnego modelu systemu.
- Przygotowanie listy metod/funkcji i klas, które potrzebne będą do realizacji zadania (jeszcze bez szczegółów kodu). Przygotowanie krótkiego komentarza do kodu. Dokumentowanie argumentów wejściowych i zwracanych wartości. Komentowanie zmiennych jest przydatne, ma jednak charakter drugorzędny.
- Implementacja programu dla prostego przypadku wejściowego z zastosowaniem wybranego paradygmatu (niekoniecznie obiektowy).
- Analiza działania programu dla danych docelowych, które stosowane będą w eksperymencie.
- * Opracowanie modelu ogólnego działającego dla dowolnych danych (w szczególności dla zmiennej liczby parametrów, uwzględniającego wiedzę dziedzinową).

- Ustalenie formatu danych wejściowych. Prezentacja wyników niekoniecznie odbywa się w bieżącym programie. Nierzadko wyniki pośrednie w programie analizowane są przez program zewnętrzny (obliczenia rozproszone, obliczenia równoległe, wielokryterialne wspomaganie decyzji, metody interaktywne).
- Wstępna analiza wyników i sprawdzenie, jakie aspekty powinny być przebadane. Ten element przeważnie pojawia się na początku, jeszcze przed tworzeniem modelu.
- Wybór reprezentacji wyników (wykres, tabela, wykres interaktywny, badanie wybranego aspektu algorytmu).
- Dokładny opis danych wejściowych z podaniem źródeł (zakresu dat dla szeregów czasowych), opisanie procedury preprocessingu danych w taki sposób, żeby możliwe było odtworzenie całej procedury i otrzymanie danych w postaci RAW.
- Przedstawienie i opis uzyskanych wyników. Próba wyciągnięcia wniosków zgodnie z zasadą generalizacji. Szukamy ogólnych schematów, wzorców, które pojawiły się w danych. Brak zgodności w jednym z przypadków nie wyklucza ogólnego wniosku płynącego z pozostałych danych. Może to być wskazanie szczególnego przypadku, który powinien być dokładnie przebadany.

Część eksperymentalna

- Powtórzenie badań (tylko w przypadku mechanizmów niedeterministycznych) minimum 30 razy (Kahneman i prawo małych liczb).
- * Analiza statystyczna. Test statystyczny potwierdzający, bądź odrzucający daną hipotezę - np. czy wyniki uzyskane dla różnych próbek są podobne, lub nie.
- Dla danych uzyskanych w wyniku działania mechanizmu niedeterministycznego potrzebne są podstawowe statystyki: średnia jest jedną z gorszych miar w tym przypadku (średnio z moim kotem mamy po trzy nogi). Mediana i odchylenie standardowe + kwartyle pozwalają na generowanie wykresów pudełkowych a także wskazują ewentualne outliersy na wykresie.
- * Analiza wrażliwości parametru pozwala na dobór optymalnych (zbliżonych do optymalnych) wartości parametrów - dla wybranego zestawu danych. Naszym celem jest dobór parametrów uniwersalnych, no ale:
- "No free lunch theorem" - średnia skuteczność algorytmów dla danej klasy problemów jest taka sama.