

ZASTOSOWANIE INFORMATYKI W MEDYCYNIE

LISTA ZAGADNIEŃ I WSKAZÓWEK PRZED TESTEM KOŃCOWYM

Wielokryterialne wspomaganie decyzji

Materiały: prezentacje na stronie.

1. Kategorie problemów decyzyjnych: klasyfikacja, porządkowanie, wybór.
2. Rodzaje kryteriów: kryteria typu zysk i typu koszt.
3. Relacja dominacji między wariantami: wariant a dominuje b , jeżeli jest lepszy na co najmniej jednym kryterium, a na żadnym nie jest gorszy.
4. Relacja preferencji (jeden wariant jest dla decydenta lepszy od drugiego) i nierozróżnialności (oba warianty są dla decydenta tak samo dobre).
5. Metoda UTA.
 - 5.1. Informacja preferencyjna (podawana przez użytkownika na wejście): preporządek zupełny (czyli ranking, w którym warianty mogą być ex-aequo) na zbiorze wybranych wariantów referencyjnych.
 - 5.2. Parametry: liczba złamań cząstkowych funkcji użyteczności.
 - 5.3. Zwracany wynik: cząstkowe funkcje użyteczności, które jednoznacznie określają globalną funkcję użyteczności. Globalna funkcja użyteczności służy nam do stworzenia finalnego rankingu na wszystkich wariantach.
 - 5.4. Cząstkowa funkcja użyteczności: dla danego kryterium określa jakość dla decydenta poszczególnych wartości. W UTA funkcje użyteczności cząstkowej są odcinkami liniowe, monotoniczne oraz zwracają wartości w przedziale $[0, k]$, gdzie k można interpretować jako wagę tego kryterium.
 - 5.5. Globalna funkcja użyteczności: suma wartości cząstkowych funkcji użyteczności danego wariantu na wszystkich kryteriach. Przyjmuje wartość 0 dla teoretycznie najgorszego możliwego wariantu (niekoniecznie występującego w zbiorze rozważanych wariantów) i 1 dla teoretycznie najlepszego możliwego wariantu (także niekoniecznie występującego w zbiorze rozważanych wariantów).

Przykładowe zadania: 1) Które z poniższych elementów są informacją preferencyjną podawaną przez decydenta w metodzie UTA? 2) Jak można zinterpretować sytuację, że funkcja użyteczności cząstkowej dla jakiegoś kryterium ma w całej dziedzinie wartość 0? 3) W jaki sposób konstruowana jest globalna funkcja użyteczności? 4) Czy jest sens umieścić w zbiorze wariantów referencyjnych wszystkie warianty?

Optymalizacja i programowanie liniowe (PL)

Materiały: prezentacje na stronie.

1. Co jest celem optymalizacji i jakie kluczowe pojęcia są z nią związane?
2. Ogólna wiedza na temat rozwiązywania zadania PL w arkuszu kalkulacyjnym, np. na co muszą być utworzone komórki i jaką informację powinny zawierać.
3. Programowanie liniowe i zapis zadania (funkcja celu, ograniczenia).

Przykładowe zadania: 1) Co wchodzi w skład przestrzeni rozwiązań w problemie

optymalizacyjnym? 2) Czy może się zdarzyć taka sytuacja, że nie ma żadnego dopuszczalnego rozwiązania optymalnego? 3) Czy ograniczenie „ $A*B + C > 0$ ” jest poprawnym ograniczeniem w zadaniu PL, gdzie A,B,C to zmienne decyzyjne? 4) Kupiec zapłaci 5 zł za każdy towar A i 10 za każdy towar B. Ma jednak łącznie tylko 1000 zł. Jakie ograniczenie opisze powyższą sytuację?

Podstawy uczenia maszynowego

Materiały: skrypt z uczenia maszynowego (guide_ml.pdf), sekcje: 2, 3, 4.1, 4.2, 5.

1. Dane uczące w problemach uczenia maszynowego. (przykłady uczące. atrybuty, decyzje)
2. Definicja klasyfikatora i do czego jest on wykorzystywany.
3. Uczenie nadzorowane a nienadzorowane.
4. Skale pomiarowe i różnice między nimi.
5. Metody oceny klasyfikatorów.
 - 5.1. Trafność klasyfikacji (% poprawnie zaklasyfikowanych przykładów).
 - 5.2. Wizualizacja wyników za pomocą macierzy pomyłek.
6. Co to jest zbiór uczący i zbiór testowy? Dlaczego został wprowadzony taki podział?
7. Techniki podziału na zbiór uczący i testowy:
 - 7.1. k-fold cross-validation (określamy liczbę podziałów (k) zbioru wszystkich przykładów i testujemy klasyfikator k razy, za każdym razem wybierając inny podział jako zbiór testowy i ucząc się na pozostałej części; następnie uśredniamy k otrzymanych wartości wybranej miary jakości).
 - 7.2. percentage split (określamy z góry, że $X\%$ wszystkich przykładów uczących to zbiór testowy).

Przykładowe zadania: 1) Do czego służy klasyfikator? 2) Jak zdefiniowana jest trafność klasyfikacji? 3) Na jakiej skali pomiarowej najlepiej wyrazić będzie informację o płci? 4) Czy na podstawie macierzy pomyłek można obliczyć trafność klasyfikacji? 5) Co opisuje k w k-fold cross-validation. 6) Dlaczego nie powinno się testować klasyfikatora na tym samym zbiorze, na którym się uczył?

K-NN

Materiały: skrypt z uczenia maszynowego (guide_ml.pdf), sekcja 8.

1. Główna idea – klasyfikacja na podstawie analogii/podobieństwa.
2. Jak przebiega uczenie? Faza uczenia sprowadza się do zapamiętania przykładów.
3. Jak przebiega klasyfikacja? Wybór k najbliższych sąsiadów rozważanego przykładu przy zdefiniowanej mierze odległości. Na ich podstawie podejmowana jest większościowa decyzja o przydziale do klasy.

Przykładowe zadania: 1) Zbiór uczący zawiera przykłady (0, 1, Tak), (2, 3, Nie), (4, 4, Nie), (2, 4, Tak). Jaką decyzję podjąłby 1-NN dla (4, 2) a jaką dla (2, 2)? Miarą odległości jest suma bezwzględnych różnic na obu współrzędnych. 2) Co się dzieje, jeżeli k w k-NN jest równe liczbie przypadków uczących? Jakie decyzje będą wtedy zwracane?

Drzewa decyzyjne

Materiały: skrypt z uczenia maszynowego (guide_ml.pdf), sekcja 12 (bez 12.5).

1. Główna idea – klasyfikacja na podstawie stopniowego zawężania zbioru przypadków za pomocą warunków/pytań o wartości atrybutów.
2. Budowa drzewa. Węzły wewnętrzne (z testami na atrybutach) i liście (decyzje). Korzeń (węzeł od którego rozpoczynamy). Gałęzie (z etykietami zawierającymi wartość atrybutu).
3. Jak przebiega uczenie? Chodzi o ogólną wiedzę, a nie o umiejętność liczenia. Wybór atrybutu, który dokona podziału zbioru na podzbiory o możliwie jednorodnych decyzjach. Do wyboru atrybutu wykorzystywana jest entropia, którą można traktować jako miarę nieuporządkowania.
4. Jak przebiega klasyfikacja? Przechodzenie po drzewie od korzenia do liścia zawierającego decyzję, ścieżka determinowana przez warunki na wartości atrybutów.
5. Entropia i jej liczenie dla prostych przypadków. Przyjmuje 0 gdy wszystkie elementy w zbiorze są takie same (prawdopodobieństwa 1 i 0, po obliczeniu ze wzoru wychodzi 0), 1 – elementy są pół na pół (prawdopodobieństwa 0,5 i 0,5, po obliczeniu ze wzoru wychodzi 1).

Przykładowe zadania: 1) Do czego wykorzystywana jest miara entropii podczas uczenia drzewa decyzyjnego? 2) Czy drzewo decyzyjne może składać się wyłącznie z korzenia? Dla jakich danych by to miało miejsce? 3) Kiedy podczas uczenia drzewa przerywa się proces dalszego podziału na atrybutach?