Uczenie maszynowe:
wykład 4

Paweł Cichosz

Kompleksy i selektory

$$\langle v_1, ?, v_3' \vee v_3'', v_4, ? \rangle$$
$$\Updownarrow$$
$$a_1 = v_1 \wedge a_3 \in \{v_3', v_3''\} \wedge a_4 = v_4$$

Kompleks: $\langle s_1, s_2, \ldots, s_n \rangle$ – wektor (koniunkcja) selektorów dla poszczególnych atrybutów, *spełniony* gdy spełnione wszystkie selektory.

Selektor: s_i – warunek określający zbiór wartości dozwolonych V_{s_i} atrybutu a_i, *spełniony* gdy wartość atrybutu należy do zbioru wartości dozwolonych (przy czym rozważamy tylko atrybuty dyskretne):

- pojedynczy: v – zbiór wartości dozwolonych $V_{s_i} = \{v\}$,
- dysjunkcyjny: $v_1 \vee v_2 \vee \ldots$ – zbiór wartości dozwolonych $V_{s_i} = \{v_1, v_2, \ldots\}$,
- uniwersalny: $?$ – zbiór wartości dozwolonych $V_{s_i} = A_i$,
- pusty: $\emptyset$ – zbiór wartości dozwolonych $V_{s_i} = \emptyset$.

Kompleksy i selektory

Kompleks uniwersalny: $\langle ? \rangle$ – zawsze spełniony (selektor uniwersalny dla każdego atrybutu).

Kompleks pusty: $\langle \emptyset \rangle$ – zawsze niespełniony (selektor pusty dla któregokolwiek atrybutu).

Kompleks jako model: $h(x) = 1$ wtedy i tylko wtedy, gdy kompleks reprezentujący h *jest spełniony* dla przykładu x (inaczej mówiąc, kompleks *pokrywa* x).

Przykład kompleksów: figury geometryczne

- Dziedzina figur geometrycznych z atrybutami:
  rozmiar: $X \to \{\mathit{mały}, \mathit{średni}, \mathit{duży}\}$
  kolor: $X \to \{\mathit{czerwony}, \mathit{niebieski}, \mathit{zielony}\}$
  kształt: $X \to \{\mathit{koło}, \mathit{kwadrat}, \mathit{trójkąt}\}$
- Kompleksy:

$$\langle \mathit{ma}, ?, \mathit{kw} \rangle$$
$$\langle \mathit{ma} \vee \mathit{śr}, ?, \mathit{kw} \rangle$$
$$\langle ?, \mathit{cze} \vee \mathit{nie}, \mathit{ko} \vee \mathit{kw} \rangle$$

Relacja większej/mniejszej ogólności/szczegółowości

- Model h_1 jest bardziej ogólny (mniej szczegółowy) niż model h_2 (model h_2 jest mniej ogólny/bardziej szczegółowy niż h_1), $h_1 \succ h_2$ ($h_2 \prec h_1$) wtedy i tylko wtedy gdy:

$$\{x \in X \mid h_2(x) = 1\} \subset \{x \in X \mid h_1(x) = 1\}$$

- Relacja częściowego porządku w przestrzeni modeli.
- Definicja odwołuje się do relacji zawierania się podzbiorów dziedziny, ale dla przestrzeni modeli o ustalonej reprezentacji porównywanie możliwe bez ich jawnego wyznaczania.

Przykład porównywania: prostokąty

- $h_1 \succ h_2$ jeśli prostokąt reprezentujący h_2 jest zawarty w prostokącie reprezentującym h_1.
- Modele reprezentowane przez prostokąty, z których żaden nie jest zawarty w drugim, nie są porównywalne.

Przykład porównywania: koniunkcje boolowskie

- $h_1 \succ h_2$ jeśli zbiór literałów koniunkcji reprezentującej h_1 jest zawarty w zbiorze literałów koniunkcji reprezentującej h_2.
- Modele reprezentowane przez koniunkcje, dla których między zbiorami literałów nie zachodzi relacja zawierania się, nie są porównywalne.

Przykład porównywania: kompleksy

- $h_1 \succ h_2$ jeśli $h_1 = \langle s'_1, s'_2, \dots, s'_n \rangle$, $h_2 = \langle s''_1, s''_2, \dots, s''_n \rangle$ oraz:

$$(\forall i = 1, 2, \dots, n)\; V_{s''_i} \subseteq V_{s'_i}$$
$$(\exists i = 1, 2, \dots, n)\; V_{s''_i} \subset V_{s'_i}$$

- Modele nie są porównywalne, jeśli $h_1 \not\succ h_2$ i $h_2 \not\succ h_1$.

Przykład porównywania: figury geometryczne

- Dziedzina figur geometrycznych z atrybutami:
  rozmiar: $X \to \{\mathit{mały}, \mathit{średni}, \mathit{duży}\}$
  kolor: $X \to \{\mathit{czerwony}, \mathit{niebieski}, \mathit{zielony}\}$
  kształt: $X \to \{\mathit{koło}, \mathit{kwadrat}, \mathit{trójkąt}\}$
- $\langle \mathit{ma} \vee \mathit{śr}, ?, ? \rangle \succ \langle \mathit{ma}, ?, \mathit{kw} \rangle$
- $\langle \mathit{ma} \vee \mathit{śr}, ?, \mathit{kw} \rangle$ i $\langle \mathit{ma}, ?, ? \rangle$ nieporównywalne.

Ogólne i szczegółowe ograniczenie przestrzeni wersji

Ograniczenie ogólne:

$$G = \{h \in \mathrm{VS} \mid \neg(\exists h' \in \mathrm{VS}) h' \succ h\}$$

Ograniczenie szczegółowe:

$$S = \{h \in \mathrm{VS} \mid \neg(\exists h' \in \mathrm{VS}) h' \prec h\}$$

Reprezentacja przestrzeni wersji za pomocą ograniczeń: $h \in \mathrm{VS}$ wtedy i tylko wtedy, gdy spełnione są oba poniższe warunki:

$$h \in G \vee (\exists h' \in G)\ h \prec h' \qquad \text{(w skrótowym zapisie: } h \preceq G\text{)}$$

$$h \in S \vee (\exists h' \in S)\ h \succ h' \qquad \text{(w skrótowym zapisie: } h \succeq S\text{)}$$

Algorytm eliminacji kandydatów (CAE)

1. inicjalizuj G jako zbiór maksymalnie ogólnych modeli w $\mathbb{H}$;
2. inicjalizuj S jako zbiór maksymalnie szczegółowych modeli w $\mathbb{H}$;
3. **dla wszystkich** $x \in T$:
   1. **jeżeli** $c(x) = 1$:
      1. $G := G - \{h \in G \mid h(x) = 0\}$;
      2. **dla wszystkich** $h \in \{h' \in S \mid h'(x) = 0\}$:
         $S := S - \{h\} \cup$ *generalizacja*(h, x, G);
      3. $S := S - \{h \in S \mid (\exists h' \in S)\ h' \prec h\}$;
   2. **jeżeli** $c(x) = 0$:
      1. $S := S - \{h \in S \mid h(x) = 1\}$;
      2. **dla wszystkich** $h \in \{h' \in G \mid h'(x) = 1\}$:
         $G := G - \{h\} \cup$ *specjalizacja*(h, x, S);
      3. $G := G - \{h \in G \mid (\exists h' \in G)\ h' \succ h\}$.

Generalizacja

- Stosowana do każdego modelu $h \in S$ niepokrywającego przetwarzanego przykładu pozytywnego x.
- Generuje zbiór wszystkich maksymalnie szczegółowych modeli h' spełniających warunki:
  - $h' \succ h$,
  - $h'(x) = 1$,
  - $h' \in G \vee (\exists h'' \in G) h' \prec h''$ (w skrótowym zapisie: $h' \preceq G$).
- Wynikiem generalizacji są modele wystarczająco ogólne, aby pokryć x, ale poza tym maksymalnie szczegółowe (generalizowane w minimalnym niezbędnym stopniu) i niewykraczające poza ograniczenie G.
- Realizacja zależna od reprezentacji modeli.

Specjalizacja

- Stosowana do każdego modelu $h \in G$ pokrywającego przetwarzany przykład negatywny x.
- Generuje zbiór wszystkich maksymalnie ogólnych modeli h' spełniających warunki:
  - $h' \prec h$,
  - $h'(x) = 0$,
  - $h' \in S \vee (\exists h'' \in S) h' \succ h''$ (w skrótowym zapisie: $h' \succeq S$).
- Wynikiem specjalizacji są modele wystarczająco szczegółowe, aby wykluczyć pokrywanie x, ale poza tym maksymalnie ogólne (specjalizowane w minimalnym niezbędnym stopniu) i niewykraczające poza ograniczenie S.
- Realizacja zależna od reprezentacji modeli.