Uczenie maszynowe: *wykład 5*

Paweł Cichosz

Przykład generalizacji: prostokąty

- Generalizacja przez minimalne rozszerzenie prostokąta reprezentującego h wystarczające do pokrycia przykładu x.
- Wynikiem jest jeden model reprezentowany przez prostokąt uzyskany przez minimalne przesunięcie jednego lub dwóch boków na zewnątrz.
- Dla uzyskanego prostokąta konieczna weryfikacja, czy wciąż jest zawarty w prostokącie reprezentującym pewien model z G.

Przykład generalizacji: koniunkcje boolowskie

- Generalizacja przez usunięcie tych i tylko tych literałów, które nie są spełnione dla dla przykładu x.
- Dla uzyskanej koniunkcji konieczna weryfikacja, czy zbiór jej literałów zawiera się w zbiorze literałów koniunkcji reprezentującej pewien model z G lub jest mu równy.

Przykład generalizacji: kompleksy

- Generalizacja przez minimalne wystarczające rozszerzenie zbiorów wartości dozwolonych selektorów.
- Dla $h = \langle s_1, s_2, \ldots, s_n \rangle$ wynikiem jest jeden model $h' = \langle s'_1, s'_2, \ldots, s'_n \rangle$ taki, że:

$$V_{s'_i} = V_{s_i} \cup \{a_i(x)\}$$

jeśli spełniony jest warunek $h' \preceq G$ (w przeciwnym przypadku generalizacja daje wynik pusty).

Przykład specjalizacji: prostokąty

- Specjalizacja przez minimalne zawężenie prostokąta reprezentującego h wystarczające do wykluczenia pokrywania przykładu x.
- Wynikiem jest zbiór wszystkich modeli reprezentowanych przez prostokąty uzyskane przez minimalne przesunięcie jednego z boków do wewnątrz.
- Dla każdego uzyskanego prostokąta konieczna weryfikacja, czy wciąż zawiera prostokąt reprezentujący pewien model z S.

Przykład specjalizacji: koniunkcje boolowskie

- Specjalizacja przez dodanie jednego literału niespełnionego dla przykładu x.
- Wynikiem jest zbiór wszystkich koniunkcji uzyskanych przez dodanie literału a_i jeśli $a_i(x) = 0$ albo $\neg a_i$ jeśli $a_i(x) = 1$.
- Dla uzyskanych koniunkcji konieczna weryfikacja, czy zbiór jej literałów zawiera zbiór literałów koniunkcji reprezentującej pewien model z S lub jest mu równy.

Przykład specjalizacji: kompleksy

- Specjalizacja przez minimalne wystarczające zawężenie zbioru wartości dozwolonych jednego selektora.
- Dla $h = \langle s_1, s_2, \ldots, s_n \rangle$ wynikiem jest *zbiór* wszystkich modeli $h'_i = \langle s_1, \ldots, s'_i, \ldots, s_n \rangle$ różniących się od h wyłącznie jednym selektorem s'_i, dla którego:

$$V_{s'_i} = V_{s_i} - \{a_i(x)\}$$

  jeśli modele te spełniają warunek $h' \succeq S$ (w przeciwnym przypadku model h'_i nie wchodzi w skład wynikowego zbioru modeli).
- Jeśli $V_{s_i} = \{a_i(x)\}$, to $V_{s'_i} = \emptyset$ i $h'_i = \langle \emptyset \rangle$, a więc oczywiście $h' \not\succeq S$ gdy $S \neq \{\langle \emptyset \rangle\}$.

Przykład CAE: figury geometryczne

x	rozmiar	kolor	kształt	c
1	*duży*	*czerwony*	*koło*	1
2	*mały*	*czerwony*	*kwadrat*	0
3	*mały*	*czerwony*	*koło*	1
4	*duży*	*niebieski*	*koło*	0

$G = \{\langle ? \rangle\}$, $S = \{\langle \emptyset \rangle\}$

Przykład CAE: figury geometryczne

x	rozmiar	kolor	kształt	c
1	duży	czerwony	koło	1
2	mały	czerwony	kwadrat	0
3	mały	czerwony	koło	1
4	duży	niebieski	koło	0

$G = \{\langle ? \rangle\},\ S = \{\langle \emptyset \rangle\}$

1. $G = \{\langle ? \rangle\},\ S = \{\langle \mathit{du}, \mathit{cze}, \mathit{ko} \rangle\}$
2. $G = \{\langle \mathit{śr} \vee \mathit{du}, ?, ? \rangle, \langle ?, ?, \mathit{ko} \vee \mathit{tr} \rangle\},\ S = \{\langle \mathit{du}, \mathit{cze}, \mathit{ko} \rangle\}$
3. $G = \{\langle ?, ?, \mathit{ko} \vee \mathit{tr} \rangle\},\ S = \{\langle \mathit{ma} \vee \mathit{du}, \mathit{cze}, \mathit{ko} \rangle\}$
4. $G = \{\langle ?, \mathit{cze} \vee \mathit{zie}, \mathit{ko} \vee \mathit{tr} \rangle\},\ S = \{\langle \mathit{ma} \vee \mathit{du}, \mathit{cze}, \mathit{ko} \rangle\}$

Zastosowanie przestrzeni wersji do predykcji

Użycie jednego modelu: wprowadzenie obciążenia w celu wyboru jednego spośród spójnych modeli – np. maksymalnie ogólny, maksymalnie szczegółowy, jednakowo „odległy" od ograniczenia ogólnego i szczegółowego, najprostszy syntaktycznie (sprawdzający najmniejszą liczbę atrybutów) itp.

Użycie wszystkich modeli: wyznaczenie $h(x)$ dla wszystkich modeli z przestrzeni wersji, następnie głosowanie.

Możliwość aktywnego uczenia się: jeśli uczeń może zadawać zapytania o prawdziwe klasy przykładów, to najkorzystniej zapytać o przykład pokrywany przez ok. połowę modeli z aktualnej przestrzeni wersji (poznanie jego klasy zmniejszy przestrzeń wersji o ok. połowę).

Podsumowanie uczenia się przestrzeni wersji

- Kamień milowy w rozwoju uczenia maszynowego: sformułowanie uczenia się jako przeszukiwania.
- Dwukierunkowe przeszukiwanie przestrzeni modeli (od maksymalnie ogólnych do maksymalnie szczegółowych, od maksymalnie szczegółowych do maksymalnie ogólnych) bez pomijania jakiegokolwiek modelu spójnego.
- Zbyt duża złożoność dla bogatych przestrzeni modeli (eksplozja zbiorów G i S).
- Brak modeli spójnych dla ubogich przestrzeni modeli (pusty wynik).
- Konieczne dostosowanie do wymogów praktycznych:
  - zawężenie zakresu przeszukiwania (np. tylko jeden kierunek, rozważane tylko wybrane modele),
  - tolerowanie niespójności.

Reguły

Rozszerzenie reprezentacji za pomocą kompleksów:

- zamiast ubogiej reprezentacji koniunkcyjnej bogata reprezentacja DNF (alternatywa koniunkcji),
- zamiast klasyfikacji binarnej klasyfikacja wieloklasowa.

Reguła: $\mathbf{k} \rightarrow d$.

część warunkowa: kompleks $\mathbf{k}$,

część decyzyjna: klasa d.

Pokrywanie: reguła $\mathbf{k} \rightarrow d$ pokrywa przykład x, gdy jej część warunkowa pokrywa (jest spełniona przez) przykład x, co będzie zapisywane $\mathbf{k} \rhd x$.

Zbiór przykładów pokrywanych: dla dowolnego $D \subseteq X$ i kompleksu $\mathbf{k}$ oraz zbioru kompleksów K:

$$D_{\mathbf{k}} = \{x \in D \mid \mathbf{k} \rhd x\}$$

$$D_K = \{x \in D \mid (\exists \mathbf{k} \in K)\mathbf{k} \rhd x\}$$

Zbiór reguł

Model: zbiór reguł, który może zawierać wiele reguł dla tej samej klasy (wówczas ich koniunkcyjne części warunkowe traktuje się jak połączone domyślną alternatywą):

- **nieuporządkowany:** każda reguła stosowana niezależnie od pozostałych – wymaga rozstrzygnięcia sposób wyznaczania przewidywanej klasy dla przykładów niepokrywanych przez żadną regułę lub pokrytych przez dwie lub więcej reguł o różnych klasach,
- **uporządkowany:** każda reguła stosowana tylko dla przykładów niepokrytych przez wcześniejsze reguły – ostatnia reguła określa klasę domyślną, brak możliwości konfliktu reguł o różnych klasach.

Wymiar VC: równy liczbie wszystkich możliwych różnych wektorów wartości atrybutów dla atrybutów dyskretnych – ryzyko nadmiernego dopasowania.

Schemat sekwencyjnego pokrywania

1. $\mathbb{R} := \emptyset$;
2. $R := T$;
3. **jak długo** $R \neq \emptyset$:
    1. $\mathbf{k} := \textit{kompleks}(R, T)$;
    2. $d := \textit{klasa}(\mathbf{k}, R, T)$;
    3. $\mathbb{R} := \mathbb{R} \cup \{\mathbf{k} \to d\}$;
    4. $R := R - R_{\mathbf{k}}$.

Schemat sekwencyjnego pokrywania

- Wymagana konkretyzacja:

  kompleks: znalezienie kompleksu pokrywającego możliwie wiele przykładów ze zbioru R (dotychczas niepokrytych) z możliwie dużą dokładnością (jednolitą bądź wyraźnie dominującą klasą),

  klasa: wybór klasy dla przykładów pokrywanych przez znaleziony kompleks (klasa dominująca).

- Znajdowanie kompleksu zwykle realizowane przez *specjalizację* – przeszukiwanie od maksymalnie ogólnych do maksymalnie szczegółowych.

Specjalizacja kompleksów

- Operacja podobna do wyznaczania przestrzeni wersji za pomocą algorytmu CAE, lecz:
  - wprowadzone obciążenie: preferencja dla kompleksów maksymalnie ogólnych,
  - wyznaczane tylko ograniczenie ogólne (zbiór G),
  - poszukiwane maksymalnie ogólne kompleksy pokrywające wyłącznie lub w większości przykłady jednej klasy,
  - eliminowanie pokrywania przykładów innych klas przez stosowanie operacji specjalizacji.
- W celu ograniczenia złożoności zwykle zbiór G przycinany do ograniczonego podzbioru najlepszych kompleksów.
- Najlepszy znaleziony kompleks wykorzystywany do utworzenia reguły.
- Niezbędna funkcja oceny jakości.
- Różne szczegółowe algorytmy.