

Uczenie maszynowe: *wykład 6*

Paweł Cichosz

1 Indukcja reguł przez sekwencyjne pokrywanie (c.d.)

Specjalizacja AQ (uporządkowany zbiór reguł)

- 1 $G := \{\langle ? \rangle\}$;
- 2 wybierz $x_s \in R$;
- 3 $R^{(1)} := R_{c=c(x_s)}$; $R^{(0)} := R_{c \neq c(x_s)}$;
- 4 **jak długo** $R_G^{(0)} \neq \emptyset$:
 - 1 wybierz $x_n \in R_G^{(0)}$;
 - 2 **dla wszystkich** $\mathbf{k} \in \{\mathbf{k}' \in G \mid \mathbf{k}' \triangleright x_n\}$:
 $G := G - \{\mathbf{k}\} \cup \text{specjalizacja}(\mathbf{k}, x_n, x_s)$;
 - 3 $G := G - \{\mathbf{k} \in G \mid (\exists \mathbf{k}' \in G) \mathbf{k}' \succ \mathbf{k}\}$;
 - 4 $G := \text{Arg}^m \max_{\mathbf{k} \in G} \nu_{R^{(1)}, R^{(0)}}(\mathbf{k})$;
- 5 **zwróć** $\arg \max_{\mathbf{k} \in G} \nu_{R^{(1)}, R^{(0)}}(\mathbf{k})$.

Specjalizacja AQ (uporządkowany zbiór reguł)

- $\nu_{R^{(1)}, R^{(0)}}(\mathbf{k})$ – ocena jakości kompleksu \mathbf{k} .
- $\text{Arg}^m \max_{\mathbf{k} \in G} \nu_{R^{(1)}, R^{(0)}}(\mathbf{k})$ – wybór m najlepszych kompleksów z G .
- $\arg \max_{\mathbf{k} \in G} \nu_{R^{(1)}, R^{(0)}}(\mathbf{k})$ – wybór najlepszego kompleksu z G .

specjalizacja: zbiór maksymalnie ogólnych kompleksów \mathbf{k}' takich, że $\mathbf{k}' \prec \mathbf{k}$, $\mathbf{k}' \not\triangleright x_n$ i $\mathbf{k}' \triangleright x_s$ (jak dla CAE, ale zamiast warunku $\mathbf{k}' \succeq S$ jest warunek $\mathbf{k}' \triangleright x_s$).

Ocena jakości kompleksów: $\nu_{R^{(1)}, R^{(0)}}(\mathbf{k})$ – zazwyczaj na podstawie liczby pokrywanych i/lub niepokrywanych przykładów w zbiorach $R^{(1)}, R^{(0)}$.

Uporządkowanie zbioru reguł: eliminacja pokrywania przykładów klas innych niż $c(x_s)$ wyłącznie w zbiorze R (przykładów niepokrytych przez wcześniejsze reguły).

Specjalizacja AQ (równoważne tradycyjne sformułowanie)

⋮

④ jak długo $R_G^{(0)} \neq \emptyset$:

- ① wybierz $x_n \in R_G^{(0)}$;
- ② $G' := \text{dyskryminacja}(x_n, x_s)$;
- ③ $G := G \wedge G'$;

⋮

dyskryminacja: (tzw. częściowa gwiazda) zbiór maksymalnie ogólnych kompleksów \mathbf{k}' takich, że $\mathbf{k}' \not\triangleright x_n$ i $\mathbf{k}' \triangleright x_s$ (równoważne specjalizowaniu kompleksu $\langle ? \rangle$ tak, aby nie pokrywał x_s i pokrywał x_n).

Koniunkcja zbiorów kompleksów: $G \wedge G' = \{\mathbf{k} \wedge \mathbf{k}' \mid \mathbf{k} \in G, \mathbf{k}' \in G'\}$.

Przykład: pogoda

x	<i>outlook</i>	<i>temperature</i>	<i>humidity</i>	<i>wind</i>	<i>play</i>
1	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>no</i>
2	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>high</i>	<i>no</i>
3	<i>overcast</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
4	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
5	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
6	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>no</i>
7	<i>overcast</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
8	<i>sunny</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>no</i>
9	<i>sunny</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
10	<i>rainy</i>	<i>mild</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
11	<i>sunny</i>	<i>mild</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
12	<i>overcast</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>yes</i>
13	<i>overcast</i>	<i>hot</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
14	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>no</i>

Przykład: pogoda

x	<i>outlook</i>	<i>temperature</i>	<i>humidity</i>	<i>wind</i>	<i>play</i>
1	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>no</i>
2	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>high</i>	<i>no</i>
3	<i>overcast</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
4	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
5	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
6	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>no</i>
7	<i>overcast</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
8	<i>sunny</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>no</i>
9	<i>sunny</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
10	<i>rainy</i>	<i>mild</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
11	<i>sunny</i>	<i>mild</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
12	<i>overcast</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>yes</i>
13	<i>overcast</i>	<i>hot</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
14	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>no</i>

$$① R = T = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}.$$

$$② G = \{\langle ? \rangle\}.$$

③ Jako ziarno wybierany jest przykład $x_s = 1$.

$$④ R^{(1)} = \{1, 2, 6, 8, 14\}; R^{(0)} = \{3, 4, 5, 7, 9, 10, 11, 12, 13\}.$$

$$⑤ R_G^{(0)} = R^{(0)} = \{3, 4, 5, 7, 9, 10, 11, 12, 13\}.$$

Przykład: pogoda

x	<i>outlook</i>	<i>temperature</i>	<i>humidity</i>	<i>wind</i>	<i>play</i>
1	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>no</i>
2	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>high</i>	<i>no</i>
3	<i>overcast</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
4	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
5	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
6	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>no</i>
7	<i>overcast</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
8	<i>sunny</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>no</i>
9	<i>sunny</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
10	<i>rainy</i>	<i>mild</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
11	<i>sunny</i>	<i>mild</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
12	<i>overcast</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>yes</i>
13	<i>overcast</i>	<i>hot</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
14	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>no</i>

$$G = \{\langle ? \rangle\}$$

- 6 Jako ziarno negatywne wybierany jest przykład $x_n = 3$.
- 7 Jedyne kompleks w G pokrywa x_n , w wyniku jego specjalizacji dostajemy:

$$G = \{\langle \text{sunny} \vee \text{rainy}, ?, ?, ? \rangle\}$$

- 8 $R_G^{(0)} = \{4, 5, 9, 10, 11\}$.

Przykład: pogoda

x	<i>outlook</i>	<i>temperature</i>	<i>humidity</i>	<i>wind</i>	<i>play</i>
1	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>no</i>
2	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>high</i>	<i>no</i>
3	<i>overcast</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
4	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
5	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
6	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>no</i>
7	<i>overcast</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
8	<i>sunny</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>no</i>
9	<i>sunny</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
10	<i>rainy</i>	<i>mild</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
11	<i>sunny</i>	<i>mild</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
12	<i>overcast</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>yes</i>
13	<i>overcast</i>	<i>hot</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
14	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>no</i>

$$G = \{\langle \text{sunny} \vee \text{rainy}, ?, ?, ? \rangle\}$$

- 9 Jako ziarno negatywne wybierany jest przykład $x_n = 4$.
- 10 Jedyne kompleks w G pokrywa x_n , w wyniku jego specjalizacji dostajemy:

$$G = \{\langle \text{sunny}, ?, ?, ? \rangle, \langle \text{sunny} \vee \text{rainy}, \text{hot} \vee \text{cold}, ?, ? \rangle\}$$

- 11 $R_G^{(0)} = \{5, 9, 10, 11\}$.

Przykład: pogoda

x	<i>outlook</i>	<i>temperature</i>	<i>humidity</i>	<i>wind</i>	<i>play</i>
1	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>no</i>
2	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>high</i>	<i>no</i>
3	<i>overcast</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
4	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
5	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
6	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>no</i>
7	<i>overcast</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
8	<i>sunny</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>no</i>
9	<i>sunny</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
10	<i>rainy</i>	<i>mild</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
11	<i>sunny</i>	<i>mild</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
12	<i>overcast</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>yes</i>
13	<i>overcast</i>	<i>hot</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
14	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>no</i>

$$G = \{ \langle \textit{sunny}, ?, ?, ? \rangle, \\ \langle \textit{sunny} \vee \textit{rainy}, \textit{hot} \vee \textit{cold}, ?, ? \rangle \}$$

- 12 Jako ziarno negatywne wybierany jest przykład $x_n = 5$.
- 13 Drugi kompleks w G pokrywa x_n , w wyniku jego specjalizacji dostajemy:

$$G = \{ \langle \textit{sunny}, ?, ?, ? \rangle, \langle \textit{sunny}, \textit{hot} \vee \textit{cold}, ?, ? \rangle, \\ \langle \textit{sunny} \vee \textit{rainy}, \textit{hot}, ?, ? \rangle, \langle \textit{sunny} \vee \textit{rainy}, \textit{hot} \vee \textit{cold}, \textit{high}, ? \rangle \}$$

- 14 Drugi kompleks nie jest maksymalnie ogólny, po jego usunięciu:

$$G = \{ \langle \textit{sunny}, ?, ?, ? \rangle, \langle \textit{sunny} \vee \textit{rainy}, \textit{hot}, ?, ? \rangle, \langle \textit{sunny} \vee \textit{rainy}, \textit{hot} \vee \textit{cold}, \textit{high}, ? \rangle \}$$

Przykład: pogoda

x	<i>outlook</i>	<i>temperature</i>	<i>humidity</i>	<i>wind</i>	<i>play</i>
1	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>no</i>
2	<i>sunny</i>	<i>hot</i>	<i>high</i>	<i>high</i>	<i>no</i>
3	<i>overcast</i>	<i>hot</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
4	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>yes</i>
5	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
6	<i>rainy</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>no</i>
7	<i>overcast</i>	<i>cold</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
8	<i>sunny</i>	<i>mild</i>	<i>high</i>	<i>normal</i>	<i>no</i>
9	<i>sunny</i>	<i>cold</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
10	<i>rainy</i>	<i>mild</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
11	<i>sunny</i>	<i>mild</i>	<i>normal</i>	<i>high</i>	<i>yes</i>
12	<i>overcast</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>yes</i>
13	<i>overcast</i>	<i>hot</i>	<i>normal</i>	<i>normal</i>	<i>yes</i>
14	<i>rainy</i>	<i>mild</i>	<i>high</i>	<i>high</i>	<i>no</i>

$$G = \{ \langle \textit{sunny}, ?, ?, ? \rangle, \\ \langle \textit{sunny} \vee \textit{rainy}, \textit{hot}, ?, ? \rangle, \\ \langle \textit{sunny} \vee \textit{rainy}, \textit{hot} \vee \textit{cold}, \textit{high}, ? \rangle \}$$

- 15 Aby zawęzić G do $m = 2$ najlepszych kompleksów, należy wyznaczyć ocenę ich jakości – np. jako sumę:
- liczby pokrywanych przykładów klasy zgodnej z klasą ziarna,
 - liczby niepokrywanych przykładów klasy innej od klasy ziarna.
- 16 Następnie specjalizacja jest kontynuowana w podobny sposób (dokończenie przykładu pozostaje jako ćwiczenie).

Specjalizacja CN2

- Proces specjalizacji bez użycia wybranych pojedynczych przykładów (ziaren).
- Każdy kompleks specjalizowany na wszystkie możliwe sposoby modyfikujące *dokładnie jeden selektor*:
jako nowe zbiory wartości dozwolonych brane pod uwagę wszystkie podzbiory obecnego zbioru wartości dozwolonych selektora.
- Ocena jakości na podstawie rozkładu klas w zbiorze R_k (premiowana dominacja jednej klasy).

Specjalizacja CN2

- Przechowywany i aktualizowany najlepszy dotychczas statystycznie istotny kompleks:
 - stosowany test statystycznie istotnej zależności między pokryciem/niepokryciem przez kompleks a klasą przykładu (np. test χ^2 , test G),
 - statystyczna istotność wymaga wyraźnie różnego rozkładu klas wśród przykładów pokrywanych i niepokrywanych oraz wystarczająco dużej liczby przykładów pokrywanych.
- Do dalszej specjalizacji przechodzi m najlepszych kompleksów z aktualnej iteracji.
- Koniec specjalizacji po osiągnięciu maksymalnej szczegółowości.

Przykład specjalizacji CN2: figury geometryczne

$$G = \{\dots, \langle ma \vee \acute{s}r, ?, ko \vee kw \rangle, \dots\}$$

$$\Downarrow$$

$$G' = \{\dots, \langle \acute{s}r, ?, ko \vee kw \rangle, \langle ma, ?, ko \vee kw \rangle, \\ \langle ma \vee \acute{s}r, cze \vee zie, ko \vee kw \rangle, \langle ma \vee \acute{s}r, zie \vee nie, ko \vee kw \rangle, \\ \langle ma \vee \acute{s}r, cze \vee nie, ko \vee kw \rangle, \\ \langle ma \vee \acute{s}r, cze, ko \vee kw \rangle, \langle ma \vee \acute{s}r, zie, ko \vee kw \rangle, \\ \langle ma \vee \acute{s}r, nie, ko \vee kw \rangle, \\ \langle ma \vee \acute{s}r, ?, ko \rangle, \langle ma \vee \acute{s}r, ?, kw \rangle, \dots\}$$

Specjalizacja FOIL

- FOIL (*first-order inductive logic*) – algorytm do indukcji reguł w rachunku predykatów, ale tutaj dostosowany do zwykłych (zdaniowych) reguł.
- Proces specjalizacji bez użycia wybranych pojedynczych przykładów (ziaren).
- Każdy kompleks specjalizowany na wszystkie możliwe sposoby modyfikujące *dokładnie jeden selektor uniwersalny*:
jako nowe zbiory wartości dozwolonych brane pod uwagę wszystkie podzbiory zbioru wartości odpowiedniego atrybutu
- Ocena jakości na podstawie rozkładu klas w zbiorze R_k (premiowana dominacja jednej klasy) z uwzględnieniem pokrycia (premiowane pokrywanie wielu przykładów).
- W każdej iteracji pozostawiany jeden najlepszy kompleks.
- Koniec specjalizacji gdy nie poprawia się jakość.

Przykład specjalizacji FOIL: figury geometryczne

$$\langle ma \vee \acute{s}r, ?, ko \vee kw \rangle$$

$$\Downarrow$$

$$\langle ma \vee \acute{s}r, cze \vee zie, ko \vee kw \rangle,$$

$$\langle ma \vee \acute{s}r, zie \vee nie, ko \vee kw \rangle,$$

$$\langle ma \vee \acute{s}r, cze \vee nie, ko \vee kw \rangle,$$

$$\langle ma \vee \acute{s}r, cze, ko \vee kw \rangle,$$

$$\langle ma \vee \acute{s}r, zie, ko \vee kw \rangle,$$

$$\langle ma \vee \acute{s}r, nie, ko \vee kw \rangle$$

Ocena jakości kompleksów

Pokrycie: na podstawie liczby pokrywanych/niepokrywanych przykładów właściwej/niewłaściwej klasy, np.:

$$\nu_{R^{(1)}, R^{(0)}}(\mathbf{k}) = |R_{\mathbf{k}}^{(1)}| + |R^{(0)} - R_{\mathbf{k}}^{(0)}|$$

Dokładność: na podstawie ilorazu liczby pokrywanych przykładów właściwej i liczby wszystkich pokrywanych przykładów, np.:

$$\nu_{R^{(1)}, R^{(0)}}(\mathbf{k}) = \frac{|R_{\mathbf{k}}^{(1)}| + 1}{|R_{\mathbf{k}}^{(1)} \cup R_{\mathbf{k}}^{(0)}| + |C|}$$

Dominacja klasy z uwzględnieniem pokrycia: np.:

$$\nu_{R^{(1)}, R^{(0)}}(\mathbf{k}) = |R_{\mathbf{k}}^{(1)}| \log \frac{|R_{\mathbf{k}}^{(1)}|}{|R_{\mathbf{k}}^{(1)} \cup R_{\mathbf{k}}^{(0)}|}$$

Dla nieuporządkowanego zbioru reguł: ocena jakości na $T_{\mathbf{k}}$.

Przycinanie zbiorów reguł

Motywacja: zbiór reguł dokładnie dopasowany do zbioru trenującego może być nadmiernie dopasowany, lecz trudno określić odpowiednie kryterium stopu specjalizacji.

Operatory:

generalizacja reguły: zastąpienie wybranego selektora selektorem uniwersalnym,

eliminacja reguły: usunięcie wybranej reguły ze zbioru reguł.

Kryteria: ocena wpływu zastosowania operatora przycinania na błąd rzeczywisty zbioru reguł (np. estymowany na oddzielnym zbiorze przykładów).