# Wstęp do teorii wyuczalności

Nina Gierasimczuk

Forum Kognitywistyczne
20 stycznia 2007

# Plan

1. Postawienie problemu
2. Model identyfikacji w granicy
   - Rodzaje danych
   - Proces zgadywania
   - Wyniki Golda
3. Wyuczalność z danych pozytywnych
   - Podstawowe definicje
   - Własności klas gramatyk a (nie-) wyuczalność
   - Własności funkcji uczącej się
   - Twierdzenie o skończonej elastyczności

# Plan

*(. . . ) dany schemat gramatyki określa klasy możliwych hipotez; metoda interpretacji umożliwia testowanie każdej hipotezy na danych wejścia; miara ewaluacyjna wybiera najwyżej ocenianą gramatykę zgodną z danymi. Od momentu, gdy hipoteza (. . . ) zostanie wybrana, uczący się zna język przez nią definiowany. W szczególności (. . . ) jest on w stanie generować nieskończony zbiór zdań z którymi nigdy się nie zetknął.*

Noam Chomsky “Aspects of the Theory of Syntax”

# Plan

- Wyuczalność dotyczy *klas* języków.
- Gold rozważa następujące klasy języków: skończonych, regularnych, bezkontekstowych, kontekstowych, pierwotnie rekurencyjnych, rekurencyjnych i rekurencyjnie przeliczalnych.

**CEL**

Wybieramy *dowolny* język z zadanej klasy. Uczeń ma odszukać nazwę (gramatykę) opisującą ten język.

**Definicja**

*TEKST dla L jest $\omega$ – ciągiem I słów $\alpha_1, \alpha_2, \ldots \in L$, takim, że każde słowo $\alpha \in L$ pojawia się przynajmniej raz w I.*

**Definicja**

*INFORMANT dla L jest $\omega$ – ciągiem I elementów z $(A^* \times \{0,1\})$, takim, że dla każdego $\alpha \in A^*$:*

$$\begin{array}{ll} (\alpha, 1) \in I & \textit{jeśli } \alpha \in L \\ (\alpha, 0) \in I & \textit{jeśli } \alpha \notin L. \end{array}$$

- Elementy ciągu treningowego są sukcesywnie prezentowane uczniowi.
- Każdorazowo, gdy podawany jest kolejny element ciągu, uczeń na podstawie dotychczas zgromadzonych danych zgaduje nazwę nieznanego języka.
- Język L jest identyfikowany w granicy, jeśli po jakimś czasie zgadnięcia są wciąż takie same i poprawne.

| | |
|---|---|
| Pierwotnie rekurencyjny tekst z generatorem | rekurencyjnie przeliczalne<br>rekurencyjne |
| Informant | pierwotnie rekurencyjne<br>kontekstowe<br>bezkontekstowe<br>regularne<br>nadskończona |
| Tekst | języki skończone |

TABELA: Wyniki Golda

# Plan

1. Postawienie problemu
2. Model identyfikacji w granicy
   - Rodzaje danych
   - Proces zgadywania
   - Wyniki Golda
3. Wyuczalność z danych pozytywnych
   - Podstawowe definicje
   - Własności klas gramatyk a (nie-) wyuczalność
   - Własności funkcji uczącej się
   - Twierdzenie o skończonej elastyczności

# Problem z danymi pozytywnymi

- Negatywny wynik Golda o użyteczności danych pozytywnych.
- Brak *systematycznej* informacji negatywnej w naturalnym procesie przyswajania języka.
- Dyskusja: Bohannon & Stanowicz vs. Wexler & Cullicover.

# Renesans teorii wyuczalności

- Istnieją klasy języków (przecinające hierarchię Chomsky'ego) wyuczalne z danych jedynie pozytywnych (Angluin 1980).
- Dowolne ograniczenie liczby reguł w gramatykach kontekstowych sprawia, że stają się one pozytywnie wyuczalne (Shinohara 1990).
- Dostrzeżono możliwość wykorzystania informacji negatywnej w modelowaniu procesu przyswajania języka - kontrprzykłady (Angluin 1987).

# System gramatyczny (*grammar system*)

**Definicja**

*Niech $\Sigma$ będzie pewnym skończonym alfabetem. System gramatyczny to $\langle \Omega, \Theta, \mathbb{L} \rangle$, gdzie:*

- $\Omega$ – *zbiór gramatyk.*
- $\Theta$ – *zbiór zdań.*
- $\mathbb{L} : \Omega \to \mathcal{P}(\Theta)$ – *funkcja nazywająca.*
  1. *Niech $G \in \Omega$. Wtedy $\mathbb{L}(G)$ nazywamy językiem generowanym z $G$.*
  2. *Pytanie "$s \in \mathbb{L}(G)$?", dla pewnych $s \in \Theta$ oraz $G \in \Omega$, nazywamy uniwersalnym problemem należenia.*

# Funkcja ucząca się (*learning function*)

**Definicja**

*Niech $\langle \Omega, \Theta, \mathbb{L} \rangle$ będzie systemem gramatycznym. Funkcja ucząca się to (częściowa) funkcja $\varphi$, taka że:*

$$\varphi : \bigcup_{k \geqslant 1} \Theta^k \rightharpoonup \Omega.$$

Algorytm uczący się = Obliczalna funkcja ucząca się.

# Funkcja ucząca się (*learning function*)

Niech $\langle s_i \rangle_{i \in \mathbb{N}} = \langle s_0, s_1, s_2, \ldots \rangle$ — nieskończony ciąg zdań z $\Theta$. Danemu ciągowi $\langle s_j \rangle_{0 \leqslant j \leqslant k}$ funkcja $\varphi$ (o ile jest zdefiniowana) przyporządkowuje gramatykę $G_k = \varphi(\langle s_0, \ldots, s_k \rangle)$.

**Definicja**

*Mówimy, że $\varphi$ zbiega do $G$ na $\langle s_i \rangle_{i \in \mathbb{N}}$, jeśli dla wszystkich prócz skończenie wielu $k \in \mathbb{N}$, $\varphi(\langle s_0, \ldots, s_k \rangle)$ jest zdefiniowane i równe $G$.*

# Funkcja ucząca się (*learning function*)

### Definicja

*Niech $\langle \Omega, \Theta, \mathbb{L} \rangle$ będzie systemem gramatycznym i niech $\mathcal{G} \subseteq \Omega$. Funkcja $\varphi$ uczy się $\mathcal{G}$ jeśli:*

- *dla każdego języka $L \in \mathbb{L}(\mathcal{G})$,*
- *dla każdego nieskończonego ciągu $\langle s_i \rangle_{i \in \mathbb{N}}$, który wylicza elementy $L$,*
- *istnieje $G \in \mathcal{G}$, taka że $\mathbb{L}(G) = L$ oraz*
- *$\varphi$ zbiega do $G$ na $\langle s_i \rangle_{i \in \mathbb{N}}$.*

### Definicja

*Klasę gramatyk $\mathcal{G}$ nazywamy (efektywnie) wyuczalną, jeśli istnieje (obliczalna) funkcja, która uczy się $\mathcal{G}$.*

# Twierdzenie Golda

**Twierdzenie (Gold 1967)**

*W dowolnym systemie gramatycznym, klasa gramatyk $\mathcal{G}$ jest niewyuczalna, jeśli $\mathbb{L}(\mathcal{G})$ zawiera wszystkie języki skończone oraz przynajmniej jeden język nieskończony.*

# Lemat o ciągu zamykającym

**Lemat (Blum & Blum 1975)**

*Niech $\varphi$ zbiega na wszystkich nieskończonych ciągach wyliczających $L$. Wtedy istnieje skończony ciąg $\langle w_0, \ldots, w_l \rangle$ o następujących własnościach:*

- $\{w_0, \ldots, w_l\} \subseteq L$
- *dla każdego skończonego ciągu* $\langle v_0, \ldots, v_m \rangle$, *jeśli* $\{v_0, \ldots, v_m\} \subseteq L$, *to* $\varphi(\langle w_0, \ldots, w_l \rangle) = \varphi(\langle w_0, \ldots, w_l, v_0, \ldots, v_m \rangle)$.

# Istnienie punktu granicznego

**Definicja**

*Klasa języków $\mathcal{L}$ ma punkt graniczny jeśli istnieje nieskończony ciąg języków $\langle L_n \rangle_{n \in \mathbb{N}}$ w $\mathcal{L}$ taki, że:*

$$L_0 \subset L_1 \subset \ldots \subset L_n \subset \ldots$$

*oraz istnieje język $L \in \mathcal{L}$ taki, że*

$$L = \bigcup_{n \in \mathbb{N}} L_n.$$

*Język $L$ nazywamy punktem granicznym $\mathcal{L}$*

# Istnienie punktu granicznego

**Twierdzenie (Kapur 1991)**

*Jeśli $\mathbb{L}(\mathcal{G})$ ma punkt graniczny, to $\mathcal{G}$ nie jest wyuczalna.*

# Ważne ograniczenie (#)

Niech $\langle \Omega, \Theta, \mathbb{L} \rangle$ będzie systemem gramatycznym takim, że:

- Istnieje algorytm, który dla danych $s \in \Theta$ oraz $G \in \Omega$, rozstrzyga czy $s \in \mathbb{L}(G)$.
- $\mathcal{G}$ jest rekurencyjnie przeliczalnym podzbiorem $\Omega$.

# Skarżypyta

## Definicja

*Indeksowana rodzina języków niepustych $L_1, L_2, L_3, \ldots$ posiada „skarżypytę” wtw istnieje efektywna procedura, która mając na wejściu $i \geqslant 1$ wylicza zbiór ciągów $T_i$ o następujących własnościach:*

1. *$T_i$ jest skończony.*
2. *$T_i \subseteq L_i$.*
3. *Dla każdego $j \geqslant 1$, jeśli $T_i \subseteq L_j$, to $L_j$ nie jest właściwym podzbiorem $L_i$.*

## Twierdzenie (Angluin 1980)

*Indeksowana rodzina niepustych języków rekurencyjnych jest wyuczalna z danych pozytywnych wtw posiada skarżypytę.*

# Skończona grubość (*finite thickness*)

**Definicja**

*Indeksowana rodzina języków niepustych $L_1, L_2, L_3, \ldots$ ma skończoną grubość wtw dla każdego skończonego niepustego zbioru $S \subseteq \Sigma^*$, zbiór $C(S) = \{L | S \subseteq L \text{ oraz } L = L_i, \text{ dla pewnego } i\}$ jest skończony.*

**Twierdzenie (Angluin 1980)**

*Jeśli klasa języków ma skończoną grubość, to jest wyuczalna z danych pozytywnych.*

# Przykład – pattern languages

## Definicja

*Niech $\Sigma$ – skończony alfabet, $Var$ – przeliczalny zbiór zmiennych. $\Sigma \cap Var = \emptyset$.*
***Wzorzec*** *nad $\Sigma$ jest dowolnym elementem z $(\Sigma \cup Var)^+$. Niech $Pat$ – zbiór wzorców nad $\Sigma$. Dla każdego $p \in Pat$, niech $L(p)$ – zbiór ciągów, które mogą być otrzymane z $p$ poprzez konsekwentne podstawienie za każdą zmienną $x$ pewnego ciągu $w \in \Sigma^+$.*

## Przykład

Niech $\Sigma = \{0, 1, 2, \ldots, 9\}$ oraz $p = 3x_6x_642x_{37}$, wtedy $L(p) = \{3xx42y \| x, y \in \Sigma^+\}$.
Przykładowe elementy: 3004213, 31221224255, 311421.

# Skończona elastyczność

**Definicja (Nieskończona elastyczność)**

*Niech $\mathcal{L}$ będzie klasą języków. Mówimy, że $\mathcal{L}$ posiada własność nieskończonej elastyczności, jeśli istnieje nieskończony ciąg zdań $\langle s_n \rangle_{n \in \mathbb{N}}$ oraz nieskończony ciąg języków $\langle L_n \rangle_{n \in N}$ w $\mathcal{L}$ takie, że dla każdego $n \in N$:*

$$s_n \notin L_n$$

*oraz*

$$\{s_0, \ldots, s_n\} \subseteq L_{n+1}.$$

**Definicja**

*Klasa $\mathcal{L}$ posiada własność skończonej elastyczności, gdy nie posiada własności nieskończonej elastyczności.*

# Twierdzenie Wrighta

**Twierdzenie (Wright 1989)**

*Niech $\langle \Omega, \Theta, \mathbb{L} \rangle$ będzie systemem gramatycznym spełniającym "ważne ograniczenie"(#).*
*Jeśli $\mathcal{G}$ posiada skończoną elastyczność, to jest wyuczalna.*

# Ilustracja

$\mathbb{L}(\mathcal{G})$ ma punkt graniczny

$\Downarrow$

$\mathcal{G}$ jest niewyuczalna

$\Downarrow$

$\mathbb{L}(\mathcal{G})$ ma nieskończoną elastyczność

# ‘Nieszkodliwe’ ograniczenia

**Definicja**

$\varphi$ *uczy się* $\mathcal{G}$ *niezależnie od porządku (order independently), jeśli dla każdego* $L \in \mathbb{L}(\mathcal{G})$, *istnieje* $G \in \mathcal{G}$ *takie, że* $\mathbb{L}(G) = L$ *i dla każdego nieskończonego ciągu* $\langle s_i \rangle_{i \in \mathbb{N}}$, *który wylicza elementy* $L$, $\varphi$ *zbiega na* $\langle s_i \rangle_{i \in \mathbb{N}}$ *do* $G$.

**Definicja**

$\varphi$ *uczy się* $\mathcal{G}$ *dokładnie (exactly), jeśli dla każdej klasy* $\mathcal{G}'$ *takiej, że* $\varphi$ *uczy się* $\mathcal{G}'$, $\mathbb{L}(\mathcal{G}') \subseteq \mathbb{L}(\mathcal{G})$.

**Definicja**

$\varphi$ *uczy się* $\mathcal{G}$ *roztropnie (prudently), jeśli* $\varphi$ *uczy się* $\mathcal{G}$ *oraz range*$(\varphi) \subseteq \mathcal{G}$.

# ‘Nieszkodliwe’ ograniczenia

### Definicja

*$\varphi$ jest responsywna na $\mathcal{G}$, jeśli dla każdego $L \in \mathbb{L}(\mathcal{G})$ i dla każdego skończonego ciągu $\langle s_0, \ldots, s_i \rangle$ elementów $L$, $\varphi(\langle s_0, \ldots, s_i \rangle)$ jest zdefiniowana.*

### Definicja

*$\varphi$ jest spójna na $\mathcal{G}$, jeśli dla każdego $L \in \mathbb{L}(\mathcal{G})$ i dla każdego skończonego ciągu $\langle s_0, \ldots, s_i \rangle$ elementów $L$, $\varphi(\langle s_0, \ldots, s_i \rangle)$ jest albo niezdefiniowana albo $\{s_0, \ldots, s_i\} \subseteq \mathbb{L}(\varphi(\langle s_0, \ldots, s_i \rangle))$.*

# ‘Szkodliwe’ (*ergo* ciekawe) ograniczenia

### Definicja

*$\varphi$ jest ukierunkowana przez zbiór (set–driven), jeśli $\varphi(\langle s_0, \ldots, s_i \rangle)$ jest zdeterminowana przez $\{s_0, \ldots, s_i\}$. Inaczej: Niech $\{s_0, \ldots, s_i\} = \{u_0, \ldots, u_j\}$, wtedy $\varphi(\langle s_0, \ldots, s_i \rangle) \simeq \varphi(\langle u_0, \ldots, u_j \rangle)$.*

### Definicja

*$\varphi$ jest konserwatywna (conservative), jeśli dla każdego skończonego ciągu $\langle s_0, \ldots, s_i \rangle$ i dla każdego zdania $s_{i+1}$, jeśli $\varphi(\langle s_0, \ldots, s_i \rangle)$ jest zdefiniowana oraz $s_{i+1} \in \mathbb{L}(\varphi(\langle s_0, \ldots, s_i \rangle))$, to $\varphi(\langle s_0, \ldots, s_i, s_{i+1} \rangle)$ jest również zdefiniowana oraz*

$$\varphi(\langle s_0, \ldots, s_i \rangle) = \varphi(\langle s_0, \ldots, s_i, s_{i+1} \rangle).$$

# ‘Szkodliwe’ (*ergo* ciekawe) ograniczenia

**Definicja**

*Funkcję uczącą się $\varphi$ nazywamy monotoniczną rosnącą, jeśli dla wszystkich skończonych ciągów $\langle s_0, \ldots, s_n \rangle$ oraz $\langle s_{n+1}, \ldots, s_{n+m} \rangle$, jeśli $\varphi(\langle s_0, \ldots, s_n \rangle)$ oraz $\varphi(\langle s_0, \ldots, s_{n+m} \rangle)$ są zdefiniowane, to $\mathbb{L}(\varphi(\langle s_0, \ldots, s_n \rangle)) \subseteq \mathbb{L}(\varphi(\langle s_0, \ldots, s_{n+m} \rangle))$.*

**Definicja**

*Funkcję uczącą się $\varphi$ nazywamy przyrastającą (incremental), jeśli istnieje obliczalna funkcja $\psi : \Omega \times \Theta \rightharpoonup \Omega$ taka, że:*

$$\varphi(\langle s_0, \ldots, s_{n+1} \rangle) \simeq \psi(\varphi(\langle s_0, \ldots, s_n \rangle), s_{n+1}).$$

# Twierdzenie Wrighta

**Twierdzenie (Wright 1989)**

*Niech $\mathcal{L}$ oraz $\mathcal{M}$ będą klasami języków o skończonej elastyczności. Wtedy klasa $\{L \cup M | L \in \mathcal{L} \wedge M \in \mathcal{M}\}$ również posiada skończoną elastyczność.*

# Twierdzenie Kanazawy

**Definicja**

*Niech $\Sigma$ oraz $\Upsilon$ będą dowolnymi alfabetami. Relacja $R \subseteq \Sigma^* \times \Upsilon^*$ nazywana jest skończenie wartościową (finite valued), jeśli dla każdego $s \in \Sigma^*$, istnieje co najwyżej skończenie wiele $u \in \Upsilon^*$ takich, że $R(s, u)$.*

**Definicja**

*Niech $M$ będzie językiem nad $\Upsilon$. Język $R^{-1}[M]$ nad $\Sigma$ definiujemy następująco:*

$$R^{-1}[M] = \{s | \exists u (R(s, u) \wedge u \in M)\}.$$

# Twierdzenie Kanazawy

**Twierdzenie (Kanazawa 1994)**

*Niech $\mathcal{M}$ będzie klasą języków nad $\Upsilon$ o skończonej elastyczności oraz niech $R \subseteq \Sigma^* \times \Upsilon^*$ będzie relacją skoczenie wartościową. Wtedy $\mathcal{L} = \{R^{-1}[M] \| M \in \mathcal{M}\}$ posiada skończoną elastyczność.*

Bibliografię znaleźć można w streszczeniu wystąpienia, na stronie:
www.kognitywistyka.uw.edu.pl