

## Tutorial Candidate Elimination Algorithm

Candidate Elimination Algorithm dapat digunakan untuk mengambil sebuah kesimpulan yang konsisten pada semua data yang dipakai sebagai training example. Sebagai contoh di slide, kita dapat melihat contoh sebuah mobil Jepang. Di tutorial yang ini, saya akan membahas konsep dari seekor gajah yang merupakan case study di materi BinusMaya.

gray?	mammal?	large?	vegetarian?	wild?	
+	+	+	+	+	(Elephant)
+	+	+	-	+	(Elephant)
+	+	-	+	+	(Mouse)
-	+	+	+	+	(Giraffe)
+	-	+	-	+	(Dinosaur)
+	+	+	+	-	(Elephant)

Ada dua himpunan/set yang digunakan yaitu G (most general set) dan S (most specific set). Berikut langkah langkah untuk menjalankan CEA:

1. Inisiasi G dengan himpunan kosong / general
  - $G = \{\{x_1, x_2, x_3, x_4, x_5\}\}$
2. Inisiasi S dengan example positive pertama. Yang dimaksud positive adalah yang sesuai dengan yang ingin kita cari. Sebagai contoh, kita ingin membahas konsep dari gajah. Jadi example positifnya adalah yang gajah.
  - $S = \{\{\text{Gray, Mammal, Large, Vegetarian, Wild}\}\}$
3. Untuk selanjutnya, kita menerima example baik positive atau negative, lalu kita mengupdate himpunan G dan S nya sesuai dengan jenis examplenya
  - Apabila example nya positive:
    - ❖ Hilangkan himpunan di G yang tidak sesuai dengan example. Apabila isi di G berupa variable yang general, artinya itu bisa berupa bermacam macam objek.
      - Contoh:  $G = \{\{x_1, x_2, x_3, \text{Vegetarian}, x_5\}, \{x_1, \text{Mammal}, x_3, x_4, x_5\}\}$
      - Lalu examplenya: {Gray, Mammal, Large, ~Vegetarian, Wild}
      - $\{x_1, x_2, x_3, \text{Vegetarian}, x_5\}$  tidak sesuai dengan example positifnya karena example mengatakan gajah bukan Vegetarian maka himpunan tersebut dihapus sehingga G menjadi  $G = \{\{x_1, \text{Mammal}, x_3, x_4, x_5\}\}$
    - ❖ Ubah variable di S yang tidak sesuai dengan example, artinya spesifikasi di S tidak *general enough* untuk example tersebut. Maksudnya

adalah example hipotesis di S tidak sesuai dengan example positif, artinya kita harus mempersempit boundary specific menjadi lebih general. Jadi cara yang dilakukan adalah variable yang tidak sesuai itu diganti menjadi lebih general.

- Contoh:  $S = \{(Gray, Mammal, Large, Vegetarian, Wild)\}$
- Lalu examplanya:  $\{Gray, Mammal, Large, \sim Vegetarian, Wild\}$
- Dalam hal ini, variable Vegetarian tidak sesuai dengan examplanya, jadi kita mengubah variable tersebut menjadi lebih general, seperti  $S = \{(Gray, Mammal, Large, x4, Wild)\}$
- Apabila example nya negative:
  - ❖ Hilangkan himpunan di S yang sesuai dengan example.
    - Contoh:  $S = \{(Gray, Mammal, Large, x4, Wild), (Gray, x2, Large, Vegetarian, Wild)\}$
    - Lalu examplanya:  $\{\sim Gray, \sim Mammal, \sim Large, Vegetarian, \sim Wild\}$
    - $(Gray, x2, Large, Vegetarian, Wild)$  sesuai dengan example negativenya karena example mengatakan yang bukan gajah itu Vegetarian maka himpunan tersebut dihapus sehingga S menjadi  $S = \{(Gray, Mammal, Large, x4, Wild)\}$
  - ❖ Ubah variable di G yang sesuai dengan example, artinya generalisasi di G tidak *specific enough* untuk example tersebut. Maksudnya adalah, example negative masih berada di dalam boundary antara G dan S apabila example tersebut mempunyai variable yang sama dengan G. Jadi cara yang dilakukan adalah variable yang sesuai itu diganti menjadi lebih specific. Cara untuk membuat lebih specific adalah dengan mengubah himpunan di G, dengan himpunan baru yang mengandung variable-variable di S yang berbeda dengan example negative nya.
    - **Contoh 1:**  $G = \{(x1, x2, x3, x4, x5)\}$ , dan  $S = \{(Gray, Mammal, Large, Vegetarian, Wild)\}$
    - Lalu examplanya:  $\{\sim Gray, \sim Mammal, Large, \sim Vegetarian, \sim Wild\}$
    - Dalam hal ini, example berada di dalam boundary karena variable-variable di G masih bersifat general semua. Jadi kita akan mengupdate G dengan himpunan yang lebih spesifik.
    - Karena antara example dan S itu ada empat variable yang berbeda, yaitu Gray, Mammal, Vegetarian dan Wild, maka kita akan membuat

4 set di G yang berisi masing-masing variable tersebut.  $G = \{(Gray, x_2, x_3, x_4, x_5), (x_1, Mammal, x_3, x_4, x_5), (x_1, x_2, x_3, Vegetarian), (x_1, x_2, x_3, x_4, Wild)\}$

- **Contoh 2:**  $G = \{(x_1, x_2, Large, x_4, x_5)\}$ , dan  $S = \{(Gray, Mammal, Large, Vegetarian, Wild)\}$
- Lalu exemplenya:  $\{Gray, Mammal, Large, Vegetarian, \sim Wild\}$
- Dalam hal ini, example berada di dalam boundary karena variable-variable di G yang Large sama dengan variable di example yang Large. Lalu variable yang di G yang  $x_5$  masih lebih general daripada variable di example yang  $\sim Wild$ . Jadi kita akan mengupdate G dengan himpunan yang lebih spesifik.
- Karena variable large di G sama dengan variable di example negativenya (yang Large), maka kita tinggal mengupdate himpunan tersebut menjadi lebih spesifik dengan mengubah variable  $x_5$  menjadi variable Wild yang ada di S. Jadi  $G = \{(x_1, x_2, Large, x_4, Wild)\}$

4. Lakukan tes example berulang ulang sampai salah satu dari dua kondisi berikut terpenuhi
  - S dan G mempunyai himpunan yang sama (singleton) - Consistent
  - Training examples habis tapi masih ada beberapa hipotesis yang tersisa – Inconsistent

Berikut adalah jawaban untuk contoh dari gajah tersebut.

gray?	mammal?	large?	vegetarian?	wild?		
+	+	+	+	+	+	(Elephant)
+	+	+	-	+	+	(Elephant)
+	+	-	+	+	-	(Mouse)
-	+	+	+	+	-	(Giraffe)
+	-	+	-	+	-	(Dinosaur)
+	+	+	+	-	+	(Elephant)

1. Inisialisasi (Example 1 +)
  - $G = \{(x_1, x_2, x_3, x_4, x_5)\}$
  - $S = \{(Gray, Mammal, Large, Vegetarian, Wild)\}$
2. Example 2 +  $\{Gray, Mammal, Large, \sim Vegetarian, Wild\}$ 
  - $G = \{(x_1, x_2, x_3, x_4, x_5)\}$
  - $S = \{(Gray, Mammal, Large, x_4, Wild)\}$
3. Example 3 -  $\{Gray, Mammal, \sim Large, Vegetarian, Wild\}$

- $G = \{(x1, x2, Large, x4, x5)\}$
  - $S = \{(Gray, Mammal, Large, x4, Wild)\}$
4. Example 4 -  $\{\sim Gray, Mammal, Large, Vegetarian, Wild\}$
- $G = \{(Gray, x2, Large, x4, x5)\}$
  - $S = \{(Gray, Mammal, Large, x4, Wild)\}$
5. Example 5 -  $\{Gray, \sim Mammal, Large, \sim Vegetarian, Wild\}$
- $G = \{(Gray, Mammal, Large, x4, x5)\}$
  - $S = \{(Gray, Mammal, Large, x4, Wild)\}$
6. Example 6 +  $\{Gray, Mammal, Large, Vegetarian, \sim Wild\}$
- $G = \{(Gray, Mammal, Large, x4, x5)\}$
  - $S = \{(Gray, Mammal, Large, x4, x5)\}$

Dapat disimpulkan bahwa kriteria seekor gajah ada berwarna **abu-abu**, jenis **mamalia** dan **besar**.