

Learning Languages by Contextual Information

Ryo Yoshinaka

Minato Discrete Structure Manipulation System Project, Japan Science and Technology Agency

Grammatical Inference

- Algorithmic learning of formal languages
- Mathematical model of language acquisition
- Natural language processing, biological sequences etc...

Substitutable Context-Free Languages

- Clark & Eyraud (2007)
- Efficiently learnable from positive data

$$x_1y_1z_1, x_1y_2z_1, x_2y_1z_2 \in L \Rightarrow x_2y_2z_2 \in L$$

Positive data:
 the man who was hungry died .
 the man ordered dinner .
 the man died .
 the man was hungry .
 was the man hungry ?
 the man was ordering dinner .



Result:
 was the man who was hungry ordering dinner ? \Rightarrow Yes
 *was the man who hungry was ordering dinner ? \Rightarrow No

Multiple Context-Free Languages

- Mild extension of CFLs
 - Polynomial-time parsable
 - "Cross serial dependencies" in natural languages
 - "Pseudoknots" in RNA sequences
- ex. $\{ a^m b^n c^m d^n \mid n, m > 0 \}, \{ ww \mid w \in \Sigma^* \}$

Learning from positive data:

The learner conjectures a language from positive examples from the target language.

Learning from positive data and membership queries:

Besides the learner can ask the teacher whether an arbitrary sentence is in the target language.

p-dimensionally substitutable MCFLs (our work)

$$w_0x_1w_1\dots x_kw_k, w_0z_1w_1\dots z_kw_k, y_0x_1y_1\dots x_ky_k \in L$$

$$\Rightarrow y_0z_1y_1\dots z_ky_k \in L \text{ for } k \leq p.$$

ex. $\{ a^m b^n c^m d^n e^m f g^n \mid n, m \geq 0 \},$
 $\{ a_1^n b_1 c_1^n d_1 \dots a_p^n b_p c_p^n d_p \mid n \geq 0 \}$

Various kinds of MCFLs

(our work)

Including all CFLs by Clark et al. ('08) and p-dimensionally SCFLs and other MCFLs.

MCFL

k,l-substitutable CFLs (our work)

$$x_1uy_1vz_1, x_1uy_2vz_1, x_2uy_1vz_2 \in L \Rightarrow x_2uy_2vz_2 \in L$$

where $|u| = k, |v| = l.$

ex. $\{ a^n b^n \mid n > 0 \}, \{ a^n c b^n \mid n > 0 \},$
 $\{ w c w^R \mid w \in \{a,b\}^* \}, \{ w \bar{w}^R \mid w \in \Sigma^+ \}$

Various kinds of CFLs

(Clark et al. '08)

Including all k,l-substitutable CFLs, regular languages, Dyck languages, etc.

CFL

Substitutable CFLs (Clark & Eyraud '07)

Regular

k-reversible languages (Angluin '82)

$$x_1vY_1, x_1vY_2, x_2vY_1 \in L \Rightarrow x_2vY_2 \in L \text{ for } |v| = k$$

0-reversible languages (Angluin '82)

$$x_1y_1, x_1y_2, x_2y_1 \in L \Rightarrow x_2y_2 \in L$$

↑ class expansion

|| conceptual analogy