GF(2)-grammars and inherent ambiguity

Vladislav Makarov

Saint Petersburg State University

18th August, 2021

Classical and GF(2)-operations

Classical operations on formal languages

1. $K \cup L := \{ w \mid w \in K$ or $w \in L \}$
2. $K \cdot L = \{ w \mid \#$ partitions $w = uv$ with $u \in K$ and $v \in L$, is positive $\}$

✓ Boolean logic $\{\text{AND}, \text{DISJUNCTION (INCLUSIVE OR)}\} \to$ GF(2) field $\{\text{AND}, \text{EXCLUSIVE OR}\}$

GF(2)-operations (Bakinova et al., 2018)

1. $K \triangle L := \{ w \mid w \in K$ or $w \in L$, but not both $\}$
2. $K \odot L := \{ w \mid \#$ partitions $w = uv$ with $u \in K$ and $v \in L$, is odd $\}$

Properties of GF(2)-operations

Subsets of Σ^* with GF(2)-operations $\Rightarrow$ uncommutative ring:

- $K \triangle L = L \triangle K$ (commutativity of addition)
- $K \triangle \varnothing = \varnothing \triangle K = K$ ($\varnothing$ is 0)
- $K \triangle (L \triangle M) = (K \triangle L) \triangle M$ (associativity of addition)
- $K \triangle K = \varnothing$ (each element is additive inverse of itself)
- $K \odot (L \odot M) = (K \odot L) \odot M$ (associativity of multiplication)
- $(K \triangle L) \odot M = (K \odot M) \triangle (L \odot M)$ (left distributivity)
- $K \odot (L \triangle M) = (K \odot L) \triangle (K \odot M)$ (right distributivity)
- $K \odot \{\varepsilon\} = \{\varepsilon\} \odot K = K$ ($\{\varepsilon\}$ is 1)
- $\varepsilon \in K \to \exists K^{-1} \colon K \odot K^{-1} = K^{-1} \odot K = \{\varepsilon\}$
  (inverse to GF(2)-concatenation)

Ordinary and GF(2)-grammars

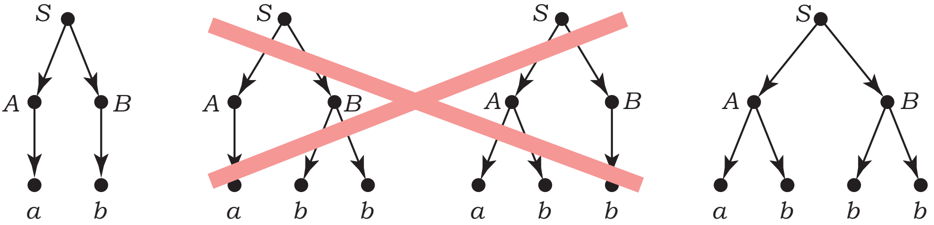
Example

$$S \to AB \qquad A \to a \mid ab \qquad B \to b \mid bb$$

$$L = \{ab, abb, abbb\}$$

$$S \to A \odot B \qquad A \to a \oplus ab \qquad B \to b \oplus bb$$

$$L' = \{ab, abbb\}$$



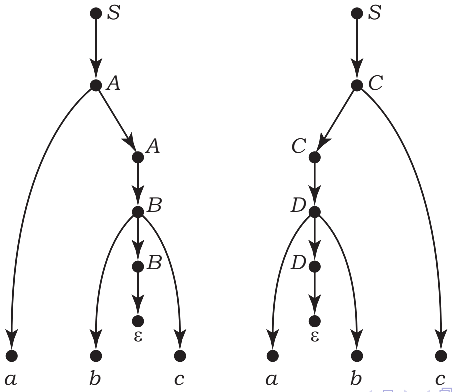
One more GF(2)-grammar example

$\{ a^\ell b^m c^n \mid \ell = m$ or $m = n$, but not both $\}$

$$S \to A \oplus C$$

$$A \to aA \oplus B \qquad C \to Cc \oplus D$$

$$B \to bBc \oplus \varepsilon \qquad D \to aDb \oplus \varepsilon$$



Language equations for ordinary and GF(2)-grammars

Ordinary grammars

- Language equations for ordinary grammars: $A = \bigcup_{A \to X_1 \dots X_\ell \in R} X_1 \dots X_\ell$
- $L(G) := \{ w \mid \#$ parse trees for w is positive $\}$.

GF(2)-grammars

- Language equations for GF(2)-grammars:
  $A = \triangle_{A \to X_1 \odot \dots \odot X_\ell \in R} X_1 \odot \dots \odot X_\ell$
- $L(G) := \{ w \mid \#$ parse trees for w is odd $\}$.

Unambiguous grammars, inherent ambiguity

Definition

Ordinary, every word has at most one parse tree $\Rightarrow$ *unambiguous.*

Most practically used grammar families are unambiguous.

Definition

No unambiguous grammar for $L \Rightarrow L$ *is inherently ambiguous.*

In general, inherent ambiguity is hard to prove.

Ogden's lemma

Theorem (Ogden's lemma in a nutshell)

$\forall p\ \exists s_1, s_2\ \exists M_1, M_2\ \forall u_1, v_1, w_1, x_1, y_1, u_2, v_2, w_2, x_2, y_2$ «*some condition*»

+ works for well-structured "small" languages
+ captures uncommutative properties pretty well
- difficult statement
- hard to use for "large" languages
- unsymmetric with respect to closure properties

Analytic approaches

Definition

For $L \subset \Sigma^*$, *its* *generating function* *is* $\mathrm{gf}(L, t) := \sum_{n=0}^{+\infty} |L \cap \Sigma^n| \cdot t^n$

Flajolet, 1987

L has an unambiguous grammar $\Rightarrow$ $\mathrm{gf}(L, z)$ is an algebraic analytic function of $z \in \mathbb{C}$, study singularities.

+ works similarly well for "very small" and "very large" languages
+ symmetric with respect to closure properties
- ignores uncommutative properties completely
- fails to work for "very simple" languages

My approach (algebraic)

1) L has an unambiguous grammar $\Rightarrow$ has a GF(2)-grammar.
2) Use strong algebraic properties of GF(2)-grammars.
3) ? (exact details vary)
4) Profit.

In particular, for subsets of $a^* b^* c^*$:

1) Language $\leftrightarrow$ power series,
2) GF(2)-grammar $\rightarrow$ linear systems over some fields,
3) language described by a grammar $\rightarrow$ element of some ring,
4) prove that L is not in the ring.

Advantages and disadvantages

+ works similarly well for “very small” and “very large” languages
+ symmetric with respect to closure properties
+ many results for the price of one.
+ captures uncommutative properties better than analytic methods
– ... but worse, than Ogden’s lemma.
– willingly ignores part of information available to analytic methods

New results

Theorem (a conjecture by Autebert et al., stated in 1979)

The language $L := \{\, a^n b^m c^\ell \mid n \neq m \;\; or \;\; m \neq \ell \,\}$ *is inherently ambiguous.*

"Large", but "simple" – bad for both Ogden's lemma and analytic methods.

Theorem (a question by Flajolet, stated in 1987)

Prove inherent ambiguity of $\{\, a^n b^m c^\ell \mid n = m \;\; or \;\; m = \ell \,\}$ *without using Ogden's lemma, preferably analytically.*

- "Small", but "simple" – good for Ogden's lemma, bad for analysis.
- Our methods handle this and the previous language at the same time.

Further directions?

Similar ideas, but in wider context?

Problem (Equivalence problem for unambiguous grammars)

Is there an algorithm that, given two unambiguous grammars, decides whether they describe the same language?