Applications of Regular Algebra to Language Theory Problems

Roland Backhouse

February 2001

Introduction

Examples:

- Path-finding problems.
- Membership problem for context-free languages.
- Error repair.

Programming Method:

- Express problem as solving a system of (recursive) equations.
- Solve the equations using eg iterative approximation or elimination technique.

Examples

$$S \ ::= \ aSS \mid \varepsilon$$

Is-empty

$$S \neq \phi \quad \equiv \quad (\{a\} \neq \phi \land S \neq \phi \land S \neq \phi) \lor \{\varepsilon\} \neq \phi$$

Nullable

$$\varepsilon \in S \quad \equiv \quad (\varepsilon \in \{a\} \land \varepsilon \in S \land \varepsilon \in S) \lor \varepsilon \in \{\varepsilon\}$$

Shortest word length

$$\#S \ = \ (\#a + \#S + \#S) \downarrow \#\varepsilon$$

Non-Example

$$S \ ::= \ aSS \mid \varepsilon$$

$$\varepsilon \in S \quad \equiv \quad (\varepsilon \in \{a\} \land \varepsilon \in S \land \varepsilon \in S) \ \lor \ \varepsilon \in \{\varepsilon\}$$

but

$$aa \in S \quad \not\equiv \quad (aa \in \{a\} \land aa \in S \land aa \in S) \ \lor \ aa \in \{\varepsilon\}$$

Problem-Solving Strategy

1. Express problem as solving a system of equations.
2. Solve equations using appropriate algorithm (iteration, elimination, Knuth's).

Constructing System of Equations

When is a function on context-free languages expressible by a system of equations with the same structure as the context-free grammar?

- "Measure" on words is extended to a "measure" on sets.
- Range of "measure" is a regular algebra.
- Measure on words is "compositional".

Theory

- Fixed Point Calculus
- Galois Connections
- Regular Algebra

Fixed Points

$$S \quad ::= \quad aSS \mid \varepsilon \quad .$$

$$S \quad = \quad \mu\langle X\!:: \{a\}\cdot X\cdot X \cup \{\varepsilon\}\rangle \quad .$$

μf denotes the least fixed point of (monotonic) endofunction f.

We sometimes write $\mu_{\leq} f$, using the subscript to indicate the ordering relation.

$\langle X\!: R\!: E\rangle$ denotes the function mapping values X in range R to E.

The range R may be omitted if it is understood from the context.

Galois Connections

Suppose $\mathcal{A} = (A, \sqsubseteq)$ and $\mathcal{B} = (B, \preceq)$ are partially ordered sets and suppose $F \in A \leftarrow B$ and $G \in B \leftarrow A$. Then (F, G) *is a Galois connection between* $\mathcal{A}$ *and* $\mathcal{B}$ iff, for all $x \in B$ and $y \in A$,

$$F.x \sqsubseteq y \ \equiv\ x \preceq G.y \quad .$$

We refer to F as the *lower adjoint* and to G as the *upper adjoint*.

Examples

Floor function:

$$n \leq \lfloor x \rfloor \ \equiv\ n \leq x \quad .$$

Negation:

$$\neg p \Rightarrow q \ \equiv\ p \Leftarrow \neg q \quad .$$

Maximum:

$$x \uparrow y \leq z \ \equiv\ x \leq z \wedge y \leq z \quad .$$

Examples (Continued)

Let $\Sigma^{\geq k}$ denote the set of all words over alphabet Σ of length at least k.

Let $\#S$ denote the length of a shortest word in the language S.

$$\#S \geq k \;\;\equiv\;\; S \subseteq \Sigma^{\geq k} \quad .$$

Fundamental Theorem

Suppose that $\mathcal{B}$ is a poset and $\mathcal{A}$ is a complete poset. Then a monotonic function $\mathsf{F}\in\mathcal{A}\leftarrow\mathcal{B}$ is a lower adjoint in a Galois connection equivales F is sup-preserving.

Examples

Let $\mathcal{S}$ denote a bag of sets. Then

$$\cup\mathcal{S} \neq \phi \;\;\equiv\;\; \exists\langle S\colon S\in\mathcal{S}\colon S\neq\phi\rangle$$

$$x\in\cup\mathcal{S} \;\;\equiv\;\; \exists\langle S\colon S\in\mathcal{S}\colon x\in S\rangle \quad .$$

Unity of Opposites

Suppose $\mathsf{F} \in \mathcal{A} \leftarrow \mathcal{B}$ and $\mathsf{G} \in \mathcal{B} \leftarrow \mathcal{A}$ are Galois connected functions, F being the lower adjoint and G being the upper adjoint. Then $\mathsf{F}.\mathcal{B}$ and $\mathsf{G}.\mathcal{A}$ are isomorphic posets.

Moreover, if one of $\mathcal{A}$ or $\mathcal{B}$ is $\mathcal{C}$-complete, for some shape poset $\mathcal{C}$, then $\mathsf{F}.\mathcal{B}$ and $\mathsf{G}.\mathcal{A}$ are also $\mathcal{C}$-complete.

Fusion Theorem

Suppose $f \in A \leftarrow B$ is the lower adjoint in a Galois connection between the complete posets $(A\,,\,\sqsubseteq)$ and $(B\,,\,\preceq)$. Suppose also that $g \in (B\,,\,\preceq) \leftarrow (B\,,\,\preceq)$ and $h \in (A\,,\,\sqsubseteq) \leftarrow (A\,,\,\sqsubseteq)$ are monotonic functions. Then

$$f.\mu_{\preceq}g \;=\; \mu_{\sqsubseteq}h \quad \Leftarrow \quad f\bullet g \;=\; h\bullet f \quad .$$

$f\bullet g$ denotes the composition of functions f and g and $f.x$ denotes application of function f to x.

An (Elementary) Application

Consider grammar

$$S \quad ::= \quad aS \;\;|\;\; SS \;\;|\;\; \varepsilon \quad .$$

We want to write $x \in S$ as a fixed point equation. That is, we want to construct g such that

$$x \in S \;\equiv\; \mu g \quad .$$

Recall:

$$x \in \cup\mathcal{S} \;\equiv\; \exists\langle S\colon S \in \mathcal{S}\colon x \in S\rangle \quad .$$

So, for all x, the boolean-valued function $(x\in)$ is a lower adjoint.

Also, $S = \mu f$ where f maps set X to

$$\{a}\cdot X \cup X\cdot X \cup \{\varepsilon\} \quad .$$

Applying fusion theorem,

$$(x \in \mu f \;\equiv\; \mu g) \;\Leftarrow\; \forall\langle S{::}\; x \in f.S \;\equiv\; g.(x \in S)\rangle \quad .$$

An Application — the empty word.

$$\varepsilon \in f.S$$

$= \quad \{$ definition of f $\}$

$$\varepsilon \in (\{a\}\cdot S \cup S\cdot S \cup \{\varepsilon\})$$

$= \quad \{$ membership distributes through set union $\}$

$$\varepsilon \in \{a\}\cdot S \;\vee\; \varepsilon \in S\cdot S \;\vee\; \varepsilon\in\{\varepsilon\}$$

$= \quad \{$ $\varepsilon \in X\cdot Y \;\equiv\; \varepsilon\in X \wedge \varepsilon\in Y$ $\}$

$$(\varepsilon\in\{a\} \wedge \varepsilon\in S) \;\vee\; (\varepsilon\in S \wedge \varepsilon\in S) \;\vee\; \varepsilon\in\{\varepsilon\}$$

$= \quad \{$ $\bullet \quad g.b \;=\; (\varepsilon\in\{a\} \wedge b) \vee (b \wedge b) \vee \varepsilon\in\{\varepsilon\}$,

see below for why the rhs has not been

simplified further $\}$

$$g.(\varepsilon\in S) \quad .$$

Thus,

$$\varepsilon \in \mu\langle X{::}\, \{a\}\cdot X \cup X\cdot X \cup \{\varepsilon\}\rangle \;\equiv\; \mu\langle b{::}\, (\varepsilon\in\{a\} \wedge b) \vee (b \wedge b) \vee \varepsilon\in\{\varepsilon\}\rangle \quad .$$

An Application — not the empty word.

$$a \in f.S$$

$=$ $\{$ definition of f $\}$

$$a \in (\{a\}{\cdot}S \cup S{\cdot}S \cup \{\varepsilon\})$$

$=$ $\{$ membership distributes through set union $\}$

$$a \in \{a\}{\cdot}S \ \vee \ a \in S{\cdot}S \ \vee \ a \in \{\varepsilon\}$$

$=$ $\{$ $a \in X{\cdot}Y \ \not\equiv \ a \in X \wedge a \in Y$ $\}$

$$??? \quad .$$

Calculation cannot be completed!!

Fusion Theorem

$$f.\mu g \;=\; \mu h \quad \Leftarrow \quad f\bullet g \;=\; h\bullet f \quad .$$

provided that

1. f is a lower adjoint
2. $f\bullet g \;=\; h\bullet f$

Strategy: f is the extension, $\widehat{m}$, to languages of a "measure" m on words. The word and language measures m and $\widehat{m}$ are constructed so that:

1. is automatically true, and
2. is true if $m.(uv) \;=\; m.u \otimes m.v \quad .$

The range of m is a regular algebra. Problem generalisation to a more sophisticated regular algebra is often needed in order to implement the strategy.

Measure $\mathfrak{m}$ on word u

$\#u \qquad$ (length of u),

true ,

$X = u$.

Extension $\hat{\mathfrak{m}}$ to language S

$\#S \;=\; \downarrow\langle u\colon u \in S\colon \#u\rangle$,

$S \neq \phi \;\equiv\; \exists\langle u\colon u \in S\colon \mathsf{true}\rangle$,

$X \in S \;\equiv\; \exists\langle u\colon u \in S\colon X = u\rangle$.

Regular Algebra

A *regular algebra* is a tuple $(A, \otimes, \oplus, \preceq, \mathbf{0}, \mathbf{1})$ where

(a) $(A, \otimes, \mathbf{1})$ is a monoid,

(b) $(A, \preceq, \oplus, \mathbf{0})$ is a complete, universally distributive lattice with least element $\mathbf{0}$ and binary supremum operator $\oplus$,

(c) for all $a{\in}A$, the endofunctions $(a\otimes)$ and $(\otimes a)$ are both lower adjoints in Galois connections between $(A, \preceq)$ and itself.

(Omitting universal distributivity, this is a Standard Kleene Algebra, Conway 1971.)

Examples

$$\mathbb{B} \;=\; (\{\mathsf{T},\mathsf{F}\}\,,\vee\,,\wedge\,,\;\Rightarrow\;,\mathsf{F}\,,\mathsf{T})\quad.$$

$$Cost \;=\; (\mathbb{R}^{\geq 0}\cup\{\infty\}\;,\;\otimes\;,\;\downarrow\;,\;\geq\;,\infty\;,0)$$

where

$$\begin{array}{lll} x{\otimes}y & = & \mathsf{if}\;\; x=\infty\vee y=\infty \;\rightarrow\; \infty \\ & & \Box\;\; x\neq\infty\wedge y\neq\infty \;\rightarrow\; x{+}y \\ & & \mathsf{fi}\;\;. \end{array}$$

$$Bottleneck \;=\; (\mathbb{R}\cup\{\infty\,,-\infty\}\;,\;\downarrow\;,\;\uparrow\;,\;\leq\;,-\infty\;,\infty)\quad.$$

$$Cost\times Bottleneck$$

where the ordering on pairs is lexicographic.

Non-Example

$$Bottleneck\times Cost\quad.$$

where the ordering on pairs is lexicographic.

Regular Homomorphism

Let $\mathcal{R} = (\mathsf{R}, \circ, \oplus, \preceq, 0_{\mathsf{R}}, 1_{\mathsf{R}})$ and $\mathcal{S} = (\mathsf{S}, \cdot, +, \leq, 0_{\mathsf{S}}, 1_{\mathsf{S}})$ be regular algebras. Suppose m is a function with domain R and range S. Then m is said to be a *regular homomorphism* if m is a monoid homomorphism (from $(\mathsf{R}, \circ, 1_{\mathsf{R}})$ to $(\mathsf{S}, \cdot, 1_{\mathsf{S}})$) and it is a lower adjoint in a Galois connection between the two orderings.

Extending Measures

Suppose that $(M\,,\cdot\,,1_M)$ is a monoid and that
$\mathcal{R} = (R\,,\circ\,,+\,,\leq\,,0_R\,,1_R)$ is a regular algebra.

Suppose m is a function with domain M and range R.

Consider the power set algebra $(2^M\,,\cdot\,,\cup\,,\subseteq\,,\phi\,,\{1_M\})$.

Define $\hat{m}$, the *extension* of m to subsets of M (elements of 2^M), by

$$\hat{m}.S \;=\; \Sigma\langle x\colon x{\in}S\colon m.x\rangle \quad .$$

Examples

$\#S$, $S\neq\phi$, $X{\in}S$.

Lemma

$\hat{m}$ is a regular homomorphism equivales m is a monoid homomorphism.

Interpreting a Grammar

Suppose $G = (N,T,P,S)$ is a context-free grammar.

Suppose $\mathcal{R} = (R\,,\times\,,\,+\,,\,\preceq\,,0_R\,,1_R)$ is a regular algebra.

Suppose m is a function with range R and domain T.

Then the *interpretation* of G in $\mathcal{R}$ under m is the endofunction on $R{\leftarrow}N$ obtained by interpreting terminal symbols via m, concatenation (on the rhs of productions) as $\times$, choice between productions as $+$, and the empty word as 1_R.

Examples

$$S \;\; ::= \;\; aSS \mid \varepsilon \;\; .$$

Interpretation of G in the regular algebra of languages under the function that maps $t{\in}T$ to $\{t\}$

$$\langle X{::}\ \{a\}{\cdot}X{\cdot}X \cup \{\varepsilon\}\rangle \;\; .$$

Interpretation of G in $\mathbb{B}$ under the function that maps t to true:

$$\langle b{::}\ (\mathsf{true} \wedge b \wedge b) \vee \mathsf{true}\rangle$$

Theorem

Suppose $G = (N,T,P,S)$ is a context-free grammar.

Suppose $\mathcal{R} = (R, \times, +, \preceq, 0_R, 1_R)$ is a regular algebra.

Suppose m is a monoid homomorphism to $(R, \times, 1_R)$ from $(T^*, \cdot, \varepsilon)$.

Suppose $\widehat{m}$ is the extension of m to the regular algebra of languages over alphabet T.

Suppose f is the interpretation of G in the regular algebra of languages under the function that maps $t{\in}T$ to $\{t\}$. Suppose g is the interpretation of G in $\mathcal{R}$ under m. Then

$$\widehat{m} \,.\, \mu f \;=\; \mu g \;\;.$$

Example — Nullability

$$\varepsilon \in \mu\langle X{::}\ \{a\}{\cdot}X \cup X{\cdot}X \cup \{\varepsilon\}\rangle \;\equiv\; \mu\langle b{::}\ (\varepsilon \in \{a\} \wedge b) \vee (b \wedge b) \vee \varepsilon \in \{\varepsilon\}\rangle$$

General CFG Recognition

Given word X and context-free grammar G, determine whether X is a word in the language generated by G.

Consider measure defined by extending m where

$$m.u \;\equiv\; X{=}u \quad .$$

This the function $(X{\in})$ on languages.

Problem: the function m is a monoid homomorphism equivales $X{=}\varepsilon$.

Solution: generalise the problem so that the range of m is a graph algebra.

Graphs

Suppose r is a binary relation and suppose A is a set. A *(labelled) graph of dimension* r *over* A is a function f with domain r and range A. Elements of relation r are called *edges*.

We will use $\mathcal{G}_{\mathsf{r}}A$ to denote the class of all labelled graphs of dimension r over A. If f is a graph and the pair (i, j) is an element of r, then $\mathsf{i}\langle \mathsf{f} \rangle \mathsf{j}$ will be used to denote the application of f to the pair (i, j).

Addition and Product of Graphs

Suppose $\mathcal{R} = (A, \times, +, \leq, 0, 1)$ is a regular algebra. Then zero and the addition and product operators of $\mathcal{R}$ can be extended to graphs as follows. Two graphs f and g of the same dimension r can be ordered according to the rule: for all pairs (i, j) in r

$$f \ \dot{\leq} \ g \ \equiv \ \forall\langle i, j{::}\ i\langle f\rangle j \leq i\langle g\rangle j\rangle \ .$$

The supremum ordering is just pointwise. In particular, f and g of the same dimension r are added according to the rule: for all pairs (i, j) in r

$$i\langle f \dot{+} g\rangle j \ = \ i\langle f\rangle j + i\langle g\rangle j \ .$$

Two graphs f and g of dimensions r and s can be multiplied to form a graph of dimension $r{\circ}s$ according to the rule: for all pairs (i, j) in $r{\circ}s$

$$i\langle f \dot{\times} g\rangle j \ = \ \sum\langle k{:}\ (i, k) \in r \wedge (k, j) \in s{:}\ i\langle f\rangle k \times k\langle g\rangle j\rangle \ .$$

Finally, the zero graph, denoted by $\mathbf{0}$, is defined by: for all pairs (i, j) in r,

$$i\langle \mathbf{0}\rangle j = 0 \ .$$

Graph Regular Algebras

Suppose $\mathcal{R} = (A\,,\times\,,+\,,\leq,0\,,1)$ is a regular algebra with carrier A, and suppose r is a reflexive, transitive relation. Define an ordering, addition and product operators as above. Define the unit graph, denoted by $\mathbf{1}$, by

$$\begin{array}{rcl} \mathsf{i}\langle\mathbf{1}\rangle\mathsf{j} & = & \textsf{if}\ \ \mathsf{i}=\mathsf{j} \rightarrow 1 \\ & & \square\ \ \mathsf{i}\neq\mathsf{j} \rightarrow 0 \\ & & \textsf{fi}\ \ . \end{array}$$

(Note that $\mathcal{G}_{\mathsf{r}}A$ is closed under the product operation and contains $\mathbf{1}$ on account of the assumptions that r is transitive and reflexive, respectively.)

Then the algebra $\mathcal{G}_{\mathsf{r}}\mathcal{R} = (\mathcal{G}_{\mathsf{r}}A\,,\dot{\times}\,,\dot{+}\,,\dot{\leq}\,,\mathbf{0}\,,\mathbf{1})$ so defined is a regular algebra.

Cocke, Kasami, Younger

Let X be a given word (the input string) and let N be the length of X.

We use X to define a measure on words and then we extend the measure to sets and then to vectors of sets. The measure of word u is a graph of Booleans that determines which segments of X are equal to u. Index the symbols of X from 0 onwards. The edge relation of the graph is the set of pairs (i,j) such that $0 \leq i \leq j \leq N$ and will be denoted by *seg*.

Now, with (i,j) in the relation *seg*, let $X[i..j)$ denote the segment of word X beginning at index i and ending at index $j{-}1$. Now define

$$m.u \;=\; \langle i,j{::}\; X[i..j) = u\rangle \quad .$$

This defines $m.u$ to be a boolean graph of dimension *seg*. Moreover, m is a monoid homomorphism and *seg* is reflexive and transitive.

$$\widehat{m}.S \;=\; \langle i,j{::}\; \exists\langle u{:}\; u{\in}S{:}\; X[i..j) = u\rangle\rangle$$

so that

$$0\langle\widehat{m}.S\rangle N \;\equiv\; X{\in}S \quad .$$

Error Repair

Let X be a given word (the input string) and let N be the length of X.

Problem: Determine the minimum number of insert, delete and/or change operations needed to edit X into a word in the language generated by context-free grammar G.

As in Cocke, Younger, Kasami, we use X to define a measure on words and then we extend the measure to sets. The measure of word u is a triangular graph of numbers that determines how many edit operations are required to transform each segment of X to the word u.

Edit Distance

Let $dist(u,v)$ denote the minimum number of non-OK edit operations needed to transform word u into word v using a sequence of the above edit operations. Now define

$$m.u \;=\; \langle i,j{::}\; dist(X[i..j)\,,u)\rangle \;\; .$$

This defines $m.u$ to be a graph of numbers. The numbers, augmented by ∞, form the min-cost regular algebra discussed earlier. Thus graphs over numbers also form a regular algebra. Taking this as the range algebra, the extension of the measure m to sets is

$$\hat{m}.S \;=\; \langle i,j{::}\; \downarrow\langle u{:}\,u{\in}S{:}\; dist(X[i..j)\,,u)\rangle\rangle$$

so that $0\langle\hat{m}.S\rangle N$ is the minimum number of edit operations required to repair the word X to a word in S.

Problem: $m.\varepsilon$ is *not* the unit of multiplication.

But, the function m does distribute through concatenation.

Compositional

Let $\mathcal{R} = (R, \circ, 1_R)$ and $\mathcal{S} = (S, \cdot, 1_S)$ be monoids. Suppose m is a function with domain R and range S. Then m is said to be *compositional* if, for all x and y in R,

$$m.(x \circ y) \;=\; m.x \cdot m.y \quad .$$

Using the Unity of Opposites

Let $\mathcal{R} = (R, \circ, \oplus, \preceq, 0_R, 1_R)$ and $\mathcal{S} = (S, \cdot, +, \leq, 0_S, 1_S)$ be regular algebras. Suppose m is a function with domain R and range S that is compositional and is a lower adjoint in a Galois connection between the orderings. Let $m.R$ be the image of R under m and let $m^\sharp$ denote its upper adjoint. Then
$m.\mathcal{R} = (m.R, \cdot, \boxplus, \leq, 0_S, m.1_R)$ is a regular algebra, where

$$x \boxplus y = m.(m^\sharp.x \oplus m^\sharp.y) \quad .$$

Moreover, m is a regular homomorphism from $\mathcal{R}$ to $m.\mathcal{R}$.

Proof

Much of the proof of this theorem is covered by the unity-of-opposites theorem — the theorem tells us that $(m.R, \leq)$ is a complete lattice with binary supremum operator $\boxplus$ as defined above and least element 0_S. To show that $m.\mathcal{R}$ is a regular algebra it thus suffices to show that $m.\mathcal{R}$ admits left and right division operators.

Proof (Continued)

Suppose $a \backslash b$ denotes right division in $\mathcal{S}$. Note that $m.x \backslash m.y$ is not necessarily in $m.R$. But,

$$
\begin{array}{ll}
 & m.y \leq m.x \backslash m.z \\
\Leftarrow & \{ \quad \text{cancellation: } m.(m^\sharp.s) \leq s \quad \} \\
 & m.y \leq m.(m^\sharp.(m.x \backslash m.z)) \\
\Leftarrow & \{ \quad \text{monotonicity of } m \quad \} \\
 & y \leq m^\sharp.(m.x \backslash m.z) \\
= & \{ \quad \text{Galois connection} \quad \} \\
 & m.y \leq m.x \backslash m.z \quad .
\end{array}
$$

Thus, in $m.R$ right division is given by the rule

$$m.x \cdot m.y \leq m.z \ \equiv \ m.y \leq m.(m^\sharp.(m.x \backslash m.z)) \quad .$$

Left division is defined similarly.

Conclusion

Heuristic for problem generalisation based on algebraic properties.