Unambiguous Recognizable Two-dimensional Languages

Marcella Anselmo, Dora Giammarresi, Maria Madonia, Antonio Restivo
(Univ. of Salerno, Univ. Roma Tor Vergata, Univ. of Catania, Univ. of Palermo)

W2DL, May 2006
REC family I

• REC family is defined in terms of local languages

• It is necessary to identify the boundary of a picture $p$ using a boundary symbol $\# \notin \Sigma$

\[
p = \begin{array}{cccc}
  & & & \\
  & & & \\
  & & & \\
  & & & \\
\end{array}
\quad \rightarrow \quad \widehat{p} = \begin{array}{cccc}
  & & & \\
  & & & \\
  & & & \\
  & & & \\
\end{array}
\]

• $L$ is local if there exists a set $\Theta$ of tiles (i.e. square pictures of size $2\times2$) such that, $p$ in $L$ if and only if any sub-picture $2\times2$ of $\widehat{p}$ is in $\Theta$
Example of local language

$L_d$ = the set of square pictures with symbol “1” in all main diagonal positions and symbol “0” in the other positions

\[ \Theta = \{ \begin{array}{ccccccccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & # & 1 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & # & 0 \\
0 & # & 0 & # & # & # & # & # & # \\
1 & # & 0 & 0 & 1 & 0 & 1 & 0 & # \end{array} \} \]

$p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \hat{p} = \begin{bmatrix} # & # & # & # & # \\ # & 1 & 0 & 0 & # \\ # & 0 & 1 & 0 & # \\ # & 0 & 0 & 1 & # \\ # & # & # & # & # \end{bmatrix} \]
**REC family II**

- **L** is recognizable by tiling system if \( L = \pi(L') \) where \( L' \) is a local language and \( \pi \) is a mapping from the alphabet of \( L' \) to the alphabet of \( L \).

**Example:** The set of all squares over \( \Sigma = \{a\} \) is recognizable by tiling system.

Set \( L' = L_d \) and \( \pi(1) = \pi(0) = a \)

\[
\begin{array}{ccc}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{array}
\in L_d \quad \pi
\begin{array}{ccc}
a & a & a \\
a & a & a \\
a & a & a \\
\end{array}
\]

- **REC** is the family of two-dimensional languages recognizable by tiling system.
About Unambiguity

• Definition of REC is implicitly non-deterministic

• The determinism and non-determinism are no more equivalent in REC: the deterministic models (4DFA, 2DOTA, ...) don’t recognize the whole REC

• REC is not closed under complement, so it is not possible to eliminate non-determinism from the model (without losing in power of recognition)

• An intermediate notion between determinism and non-determinism is the notion of unambiguity
Unambiguous Recognizable Languages

Def [GR92] A tiling system \((\Sigma, \Gamma, \theta, \pi)\) is \textit{unambiguous} for \(L \subseteq \Sigma^{**}\) if the projection \(\pi\) is injective on \(L(\theta)\) (i.e. for any \(p \in L\) there is a unique \(p' \in L'\) such that \(\pi(p') = p\)).

\(L \subseteq \Sigma^{**}\) is \textit{unambiguous} if it admits an unambiguous tiling system.

\textbf{UREC} denotes the family of all unambiguous recognizable 2dim languages.

- \(\textbf{UREC} \subseteq \textbf{REC}\)
- Generalization in 2dims of unambiguous automata for strings
Example: $L_{\text{col-1n}} = \{ p \mid \text{first } \text{col} = \text{last } \text{col} \} \subseteq \{a,b\}$

- $L_{\text{col-1n}} \in \text{REC}$

Idea: Use $\Gamma = \{x_y\}$ where

- the subscript $y$ saves the symbol of the first column and

- $\pi(x_y) = x$

\[ p = \begin{array}{ccc}
  b & b & a & b \\
  a & a & a & a \\
  b & a & b & b \\
  a & b & b & a \\
\end{array} \quad \rightarrow \quad p' = \begin{array}{ccc}
  b & b & a & b \\
  a & a & a & a \\
  b & a & b & b \\
  a & b & b & a \\
\end{array} \]

- $L_{\text{col-1n}} \in \text{UREC}$
UREC and REC

- **UREC \(\subsetneq\) REC? Yes**

\[
L_{\text{col-ij}} = \begin{cases} 
\text{col } i = \text{col } j 
\end{cases}
\]

\[
L_{\text{col-ij}} = \sum^{**} \ominus L_{\text{col-1n}} \ominus \sum^{**} \text{ and REC is closed with respect to } \ominus
\]

\[
L_{\text{col-ij}} \in \text{REC}
\]

\[
L_{\text{col-ij}} \notin \text{UREC}
\]

**WHY?**
Towards a necessary condition for unambiguity

- Reduce two dimensional languages to string languages over the alphabet of the columns (i.e. define $L(m)$)

- Use the Theorem of Hromkovic et al. for a lower bound on the states of an unambiguous automaton for a string language
From 2dim to 1dim

Let $L \subseteq \Sigma^{**}$. For any $m$ consider the subset $L(m) \subseteq L$ of all pictures with exactly $m$ rows.

- $L(m)$ can be viewed as a string language over the alphabet of the columns.

Example:

$$p = \begin{bmatrix}
    b & b & a \\
    a & a & a \\
    b & a & b \\
    a & b & b
\end{bmatrix} \in L \quad \text{the string } w = \begin{bmatrix}
    b \\
    a \\
    b \\
    a
\end{bmatrix} \begin{bmatrix}
    b \\
    a \\
    b \\
    b
\end{bmatrix} \in L(4)$$
An automaton for $L(m)$

**Theorem** [Matz 97] Let $L \subseteq \Sigma^{**}$. If $L \in \text{REC}$, then there is a $k$ such that, for all $m$, there is a finite string automaton $A_m$ with $k^m$ states for $L(m)$.

**Idea of Proof:** Let $(\Sigma, \Gamma, \theta, \pi)$ a tiling system for $L$.

- The states of $A_m$ are all the possible columns (of height $m$) in the local alphabet $\Gamma$, plus an initial state.
Idea of Proof (continued)

- There is an edge from column $p$ to column $q$ if and only if any sub-picture $2 \times 2$ of $p \oplus q$ is in $\theta$. The label for this edge is $\pi(q)$

**Example:** In $L_{\text{col-1n}}$ we have
Theorem of Hromkovic et al.

**Def** Let \( S \subseteq \Sigma^* \) be a regular string language. Define the infinite boolean matrix \( M_S = \{a_{\alpha\beta}\} \), where \( a_{\alpha\beta} = 1 \) if and only if \( \alpha\beta \in L \).

- Since \( S \) is regular, the number of different rows of \( M_S \) is finite.

Let \( S \subseteq \Sigma^* \) be a regular string language. Denote by \( \text{uns}(S) \) the size of a minimal unambiguous non-deterministic automaton accepting \( S \).

**Theorem** (Hromkovic et al.) For every regular string language \( S \subseteq \Sigma^* \), \( \text{uns}(S) \geq \text{Rank}_Q(M_S) \).
A necessary condition for unambiguity

Theorem Let $L \subseteq \Sigma^{**}$. If $L \in \text{UREC}$, then there is a $k$ such that, for all $m$, $\text{Rank}_Q(M_{L(m)}) \leq k^m$.

Proof:

• Note that if $L \in \text{UREC}$ then the automaton $A_m$ for $L(m)$ is unambiguous

• Use the inequality $\text{uns}(L(m)) \geq \text{Rank}_Q(M_{L(m)})$
Consider \( L = L_{\text{col-ij}} \)

For every \( m \), \( L(m) \) is a language of strings with at least two occurrences of the same symbol.

It is possible to show that \( M_{L(m)} \) has Rank equal to \( 2^{\left| \Sigma \right| m} + 1 \) against the necessary condition for UREC.

**Theorem (restated)** There exist recognizable 2dim languages that are inherently ambiguous.
Properties of UREC

Proposition UREC is closed under intersection and rotation operations.

Proposition UREC is not closed under row/column concatenation/closure.

Proof:

\( L_{\text{col-1n}} \in \text{UREC}. \)

But \( L_{\text{col-ij}} = \Sigma^{**} \bigodot L_{\text{col-1n}} \bigodot \Sigma^{**} \notin \text{UREC}. \)
Using automata characterization

Def A 2UOTA is a 2OTA such that it has at most one accepting run on a picture p.

Theorem \( L(2DOTA) \not\subset L(2UOTA) \not\subset L(2OTA) \).

Proof: Note that \( L(2UOTA) = \text{UREC} \) (see also Mäurer02) and \( L(2OTA) = \text{REC} \).

- For the first inclusion, consider the language \( L = \{ p \mid p \text{ is a square} \land \text{last row} = \text{last col} \} \subseteq \{a,b\}^* \)
  \( L \not\in L(2DOTA) \) but \( L \in L(2UOTA) \)

- The second inclusion follows from \( L(2UOTA) = \text{UREC} \not\subset \text{REC} = L(2OTA) \)
An undecidability result

**Theorem** Given a tiling system \((\Sigma, \Gamma, \theta, \pi)\) for \(L \subseteq \Sigma^*\), it is undecidable whether it is unambiguous.

**Proof:** By reduction from the undecidable 2dimensional Unique Decipherability Problem.
Further work

• Questions related to UREC

• Questions related to (?) DREC (deterministic version of REC)
Open Problems

• Is UREC closed under complement?

• Is UREC largest subset in REC closed under complement?

Conjecture: If $L \in \text{REC}\setminus\text{UREC}$ then $\sim L \notin \text{REC}$
About Deterministic Recognizable 2dim Languages

• Many deterministic models: 4DFA, 2DOTA, … They don’t recognize the whole REC

• In string languages the notion of determinism is, in some sense, “oriented”:
  - Determinism from left to right
  - Co-determinism from right to left

• In a picture four different directions.

• Two proposals for the definition of 2dimensional determinism.
First approach

Idea: A tiling system ($\Sigma, \Gamma, \theta, \pi$) is *Top-Left-deterministic* if $\forall a,b,c \in \Gamma$ and $s \in \Sigma \exists$ unique tile such that $\pi(s)=d$.

(Analogously *TR-, BL-, BR-deterministic* tiling system)

$L$ is *deterministic* if $L$ has a tiling system that is deterministic with respect to some direction (TL or TR or BL or BR)
Second approach

Idea: A tiling system is left-to-right column-deterministic if, after having computed the local symbols in an entire column of a picture, the local symbols on the next one are univocally determined.

L is deterministic if L has a tiling system that is deterministic with respect to one direction by column and a tiling system that is deterministic with respect to one direction by row.
Working proposal for these days

• Find an appropriate definition for determinism in terms of tiling system that is not oriented as the recognition by tiling systems
The end