CMPE 598 - Lecture 4

Scribe : Hüseyin Bilge Yağcı

Feb 27, 2018

1 Recall from last week

In last lecture, we had proved that for promise problems, QFAs are economical in number of states, compared to DFA's. Up to now, we depended on exactness on QFAs, but exactness will be disturbed in the upcoming lectures.

Today's topic What about general language recognition, without promises? In this setting, any string can appear and the machine must respond correctly. We will try to find whether QFAs have advantages - state economy or functionality - compared to DFA's in general language recognition, or not.

2 Distinguishability and Myhill-Nerode Theorem

Distinguishability. Let x and y be strings, and let L be any language. We say that x and y are distinguishable by L if some string z exists such that exactly one of the strings xz and yz is a member of L. If x and y are not distinguishable by any z, i.e. if for every z, we have $xz \in L$ iff $yz \in L$; we denote this with $x \equiv_L y$.

Distinguishability is an equivalence relation, and the set of all strings are divided into equivalence classes by L.

Index of a language. Let L be a language. We define **index** of L as the maximum number of elements in any set that are pairwise distinguishable by L.

Examples

- Let $\Sigma_1 = \{0, 1\}$, $L_1 = \{l |\ l$ ends with $1\}$. For $x = 1$, $y = 0$, $z = \epsilon$; $yz \in L_1$, but $xz \notin L_1$; thus x and y are distinguishable. In this context, $w = 110001$ and $u = 11$ are indistinguishable and $w \equiv_{L_1} u$. Here, the strings with the same ending symbol belong to the same equivalence class, i.e. L_1 has an index of 2.
- Let $\Sigma_2 = \Sigma_1, L_2 = \{l |\ l$ contains equal amounts of 1's and 0's$\}$. $x = 010$ and $y = 100$ are indistinguishable by L_2. Here, the strings are divided into classes with respect to the discrepancy between 1's and 0's inside them; therefore, index of L_2 is infinite.
- Let $\Sigma_3 = \{1\}, L_3 = \{11, 111\}$. The elements of $\{\epsilon, 1, 11, 111, 1111\}$ are pairwise distinguishable and thus form equivalence classes, but all the remaining strings are indistinguishable from 1111 by L_3.

Lemma. If L is recognizable by a DFA with k states, then it has index at most k.

Proof. Assume the index of L is greater than k, which means that there are at least $k+1$ strings that are pairwise distinguishable. Due to pidgeon-hole principle, at least two of the distinguishable strings must bring the machine to the same state. The strings become indistinguishable, therefore the assumption contradicts itself. □

Lemma. If the index of language L is a finite number k, then it is recegnized by a DFA with k states.

Proof. Let $x_i = \{S_0, S_1, \ldots, S_{k-1}\}$ be pairwise distinguishable by L. Let D be the DFA with

$$
\begin{aligned}
&D = (Q, \Sigma, \delta, q_0, F)\\
&Q = (q_0, q_1, \ldots, q_{k-1})\\
&\delta(q_i, a) \longrightarrow q_j, \text{ where } S_i a \equiv_L S_j \text{ for any } a \in \Sigma\\
&F = \{q_i | S_i \in L\}\\
&q_0 \text{ is such that } S_i \equiv_L \epsilon.
\end{aligned}
$$

Every state in D corresponds to an equivalence class, thus D recognizes L. □

Myhill - Nerode Theorem. *A language L is regular iff it has finite index.* Moreover, index of L is the size of the smallest DFA recognizing L. Proofs are omitted.

3 Communication Complexity

Communication complexity tries to quantify the minimum number of bits to be shared between two parties solving a certain problem. Let Alice and Bob be two individuals. Alice is given string x, where Bob is given y. Their aim is to figure out if string xy belongs to a certain language L. What is the smallest message that Alice can send to Bob to transfer information? The trivial solution to this is sending the full string x to Bob, but generally there exists a shorter message for this job. We will define **one-way communication complexity** of language L as the minimum number of bits that has to be sent by Alice. We will mention some concepts from information theory to solve this problem. Ultimately, our aim is to compare quantum and classical communication complexity.

3.1 Information Theoretical concepts

Shannon entropy. The Shannon entropy $\boldsymbol{H(B)}$ of a set of messages, described with random variable B coresponds to the average number of classical bits required to encode the members of this set. For a set with n members and probability distribution $p_i = \{p_1, p_2, \ldots, p_n\}$, Shannon entropy is defined as

$$\boldsymbol{H(B)} \triangleq -\sum_{x=1}^{n} p_x \log_2 p_x \tag{1}$$

Shannon entropy can be seen as the number of bits needed to represent a given set fully. From the definiton of $\boldsymbol{H(B)}$, we can see that $\boldsymbol{H(B)}$ is maximized when $\mathbf{B}$ is uniformly distributed, in which we need $\log_2 n$ bits; and zero when there is no uncertainty in $\mathbf{B}$'s outcome.

von Neumann entropy. For quantum systems, von Neumann came up with quantum version of the entropy, since the density matrix is somewhat a probability distribution. von Neumann entropy $\boldsymbol{S(\rho_B)}$ is defined as

$$\boldsymbol{S(\rho_B)} \triangleq -tr(\rho_B \log_2 \rho_B) \tag{2}$$

von Neumann entropy boils down to the number of quantum bits (qubits) needed to represent the set. $\boldsymbol{S(\rho_B)}$ is maximized when the distribution is uniform, and zero when the states are "pure", i.e. when we have complete knowledge of the system and which state it is in.

3.2 DFA communication complexity

Consider the infinite two-dimensional matrix μ, with the rows and columns represent x and y's, $x, y \in \Sigma^*$, $\Sigma = \{0, 1\}$. Define $\mu(x, y)$ such that $\mu(x, y) = 1$ iff $x, y \in L$, and 0 otherwise. Let $L = \{w | w \text{ ends with } 1\}$ for now.

$\boldsymbol{\mu}$	ϵ	0	1	00	01	10	11	...
ϵ	0	0	1	0	1	0	1	...
0	0	0	1	0	1	0	1	...
1	1	0	1	0	1	0	1	...
00	0	0	1	0	1	0	1	...
01	1	0	1	0	1	0	1	...
10	0	0	1	0	1	0	1	...
11	1	0	1	0	1	0	1	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

For regular L, as seen in this case, matrix μ has a finite number of distinct rows, corresponding to equivalence classes. Any string x that can be plugged will bring the DFA into one of the finite states. Identical rows mean *indistinguishability*. Alice's job will be only telling Bob which type of unique row x is in. The only information needed is *"X brought me to that state"*. For this particular language, we only need 1 bit because index of L is two. Similarly, for n states, we need $\log_2 n$ bits.

3.3 QFA communication complexity

Any QFA with q states can be simulated by Alice and Bob, with Alice sending the state of the QFA after processing x, which is the procedure we applied in DFA case. Alice sends $\log_2 q$ qubits to Bob and it will be sufficient. We are interested in whether the size of QFA is smaller than DFA or not. *(Spoilers: We will show that one-way quantum communication complexity of regular language L with index d is $\log_2 d$, then conclude that no QFA with fewer than d states exists for this job.)* If the index of regular language L is d, reduce the communication matrix to d distinct rows. The information to be sent to Bob is the mixed state of uniformly randomly distributed rows.

Imagine the rows being chosen randomly bit by bit, i.e. column by column. Let $p(0)$ be the probability of a 0 in the first column (i.e. $\frac{\#0\text{'s in the first column}}{d}$). Then 0 is chosen with probability $p(0)$ and 1 is chosen with $1-p(0)$. Partition the rows to the sets I_0 and I_1, the sets of rows starting with 0's and 1's, respectively. If b is chosen for the first transmitted bit, then the process continues with set I_b and the next column. If a complete row X is determined, let $\boldsymbol{\rho_X}$ denote the density matrix of just the message about that row. Let $\boldsymbol{\rho_t}$ denote the density matrix of possibly mixed message corresponding to a row, starting with t chosen uniformly among all such rows.

The probability that a b is chosen after t is called $\boldsymbol{p_t(b)}$. The associated RV is called B, and the number of different rows beginnng with t is called $\boldsymbol{row_t}$. Bob can decide membership of xy in L correctly with $p=1$, so he receives exact information about the row corresponding to X in the message sent by Alice.

Holevo Theorem. *Suppose Alice prepares a quantum state ρ_x where $x=\{0,1,\ldots,n\}$ with probability $p_i=\{p_0,p_1,\ldots,p_n\}$; and then gives it to Bob. Bob performs a measurement on that state, with measurement outcome Y.* ***The Holevo Bound*** *states that for any such measurement Bob may do,*

$$\boldsymbol{H(X:Y)\leq S(\rho)-\sum_x p_x S(\rho_x)},\ \textit{where}\ \rho=\sum_x p_x\rho_x \tag{3}$$

$H(X:Y)$ denotes mutual information between X and Y, defined as *the amount of ignorance about Y that is reduced due to knowing about X.* When X and Y are independent, they have zero mutual information, and when X and Y are identical, $H(X:Y)=H(X)=H(Y)$. In our case, we want X and Y to be identical. Because of exactness requirement,

$$S(\rho_t)\geq p_t(0)S(\rho_{t_0})+p_t(1)S(\rho_{t_1})+H(B) \tag{4}$$

for any t. We will show by induction that $S(\rho_t)\geq \log_2 row_t$.

- Basis step: $S(\rho_t)\geq 0$ for any completely chosen ρ, since von Neumann entropy is always nonnegative.
- Inductive step: Start by modifying (4).

$$S(\rho_t)\geq p_t(0)\underbrace{\log_2 row_{t_0}}_{\leq S(\rho_{t_0})}+p_t(1)\underbrace{\log_2 row_{t_1}}_{\leq S(\rho_{t_1})}+H(B) \tag{5}$$

$$\begin{aligned}\geq\ & p_t(0)\log_2\left[p_t(0)row_t\right]+p_t(1)\log_2\left[p_t(1)row_t\right]\\ & -p_t(0)\log_2 p_t(0)-p_t(1)\log_2 p_t(1)\end{aligned} \tag{6}$$

First two terms in (6) can be reorganized as $p_t(i)[\log_2[p_t(i)row_t]]\rightarrow p_t(i)[log_2 p_t(i)+\log_2 row_t]$ so that the negative contributors from $H(B)$ are negated.

$$S(\rho_t)\geq[p_t(0)+p_t(1)]\log_2 row_t=\log_2 row_t \tag{7}$$

By induction, we proved $S(\rho_t)\geq\log_2 row_t$. Selecting $t=\epsilon$, we find $S(\rho_\epsilon)\geq log_2 row_\epsilon=log_2 d$, which is same result we obtained in DFA case. Thus, we can conclude that **when the QFA is required to work with zero error, it has no state advantage over DFA for recognizing a regular language.**