CMPE 492

Entropy Hierarchy

Bilal Aytekin

Advisor:

Cem Say

June 7, 2023

TABLE OF CONTENTS

1. INTRODUCTION 1
   1.1. Broad Impact 1
2. PROJECT DEFINITION AND PLANNING 2
   2.1. Project Definition 2
   2.2. Project Planning 3
3. RELATED WORK 4
4. METHODOLOGY 5
   4.1. Observation 5
   4.2. Application 5
5. RESULTS 14
6. CONCLUSION 15
REFERENCES 16

1. INTRODUCTION

1.1. Broad Impact

Recently, in [1], the authors showed that there exists a regular language L such that every sufficiently long string "forgets" more than one bit of information at every step of its computation on every deterministic finite automaton recognizing L. In this project, we extend this result on the average behavior of information loss by constructing languages that forget half a bit, one third a bit, etc., at every step of computation, hence build a hierarchy of this so called entropy cost.

2. PROJECT DEFINITION AND PLANNING

2.1. Project Definition

An *alphabet* is a possibly infinite set of symbols. If Σ is an alphabet, Σ^* denotes the set of all possible strings in that alphabet, and $\Sigma^n \subset \Sigma^*$ those strings of length n. A *language* L on an alphabet Σ is any subset $L \subseteq \Sigma^*$.

A *deterministic finite automaton (DFA) D* is a 5-tuple $D = (\Sigma, Q, \delta, q_0, F)$ where

- Σ is a *finite* alphabet,
- Q is the set of states of D,
- $\delta \colon Q \times \Sigma \to Q$ is the transition function,
- $q_0 \in Q$ is the initial state,
- $F \subseteq Q$ is the set of accepting states of D.

Given $w = w_1 \dots w_n \in \Sigma^n$, if $(q_0, q_1, \dots, q_n)$ is the sequence of states M traverses while running on w, i.e., if $q_i = \delta(q_{i-1}, w_i)$ for all $i = 1, \dots, n$, then we say M *accepts* w if $q_n \in F$ and *rejects* otherwise. The set of all strings M accepts is the *language recognized by M*.

Example 2.1.1. The DFA $(\{a, b\}, \{q_0, q_1\}, \delta, q_0, \{q_1\})$ for

$$\delta(q_0, a) = q_0,\ \delta(q_0, b) = q_1,$$
$$\delta(q_1, a) = \delta(q_1, b) = q_0$$

is in Figure 2.1. The language recognized by this DFA is the set of all strings ending with the letter a.

△

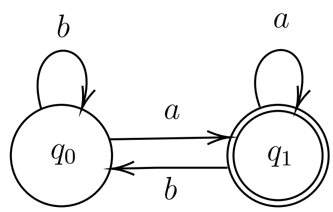

Figure 2.1. An example DFA.

Given a DFA $M = (Q, \Sigma, \delta, s, F)$, for a letter $\sigma \in \Sigma$, we denote the number of σ-transitions into state $u \in Q$ as $\chi_M(u, \sigma)$ –or $\chi(u, \sigma)$ when the referred DFA is clear.

We say a language $L \subseteq \Sigma^*$ has average entropy at most s if there is a DFA D recognizing L such that for sufficiently large n, every $w = w_1 \dots w_n \in \Sigma^n$ satisfies

$$\frac{\sum_{i=1}^{n} \chi_D(u_i, w_i)}{n} \leq s,$$

where $(u_0, u_1, \dots, u_n)$ is the sequence of states D traverses under w. Similarly, we define a language with average entropy at least s, and say a language has average entropy s if it has average entropy both at least and at most s.

Our project is to prove the following theorem:

Theorem 2.1.2 (Entropy Hierarchy)**.** *For every $n > 0$, there exists a language with entropy $1/n + \epsilon$.*

2.2. Project Planning

We have divided our time between midterms. For the first midterm, we aimed at constructing languages with entropy at most $1/n$ for every $n > 0$. In the second half of the term, we turned to dealing with regular languages according to the lower bounds on their entropy cost, which will let us conclude the proof of Theorem 2.1.2.

3. RELATED WORK

We say a computation is *reversible* if one can trace all the steps of the computation back given the state of the machine after the computation and the input it has read. In [2], Landauer showed that every bit irreversible machines "forget" causes them to dissipate heat.

In the context of DFAs, irreversibility comes from more than one incoming transitions to a state which are labeled with the same letter. Turing machines can be made irreversible by can recording which state they came from in their work tape, and, even, DFAs which have 2-way access to their input [3].

On the other hand, it is known that, reversible DFAs can recognize only a proper subset of all regular languages [4]; therefore, it is of interest to understand the change in the entropy cost over the class of regular languages [5, 6]. In particular, we are working to extend the main result about constructing a maximally expensive language of [1] to introduce languages in each level of energy expenditure.

4. METHODOLOGY

4.1. Observation

Consider the language $L = \{w \in \{a,b\}^* | \ w \text{ contains "}bb\text{" and ends with a "b"}\}$. The minimal DFA $D = (\{a,b\}, Q, \delta, 1, \{3\})$ recognizing L is given in Figure 4.1.

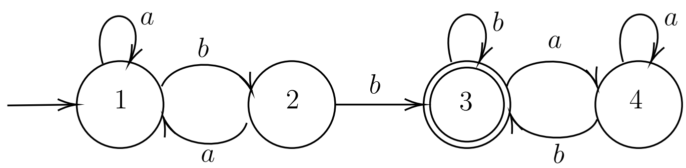

Figure 4.1. A minimal DFA D for L.

For $w \in \{a,b\}^n$, let $(u_0, u_1, \ldots, u_n)$ be the sequence of states traversed by D while running on w. Since $\chi(u,\sigma) \leq 3$ for every $u \in Q$ and $\sigma \in \{a,b\}$, we can bound the average entropy of this computation by

$$\frac{\sum_{i=1}^{n} \log_2(\chi(u_i, w_i))}{n} \leq \frac{\sum_{i=1}^{n} \log_2(3)}{n} = \log_2(3).$$

Now, we propose a scheme to cut this cost –hence the upper bound $\log_2(3)$– down by a half: Let $M = (\{a,b,x\}, Q', \delta', 1, \{3\})$ be the DFA given in Figure 4.2. We observe that adding a free[1] state ux for every original state u, and making loops of length 2 at the newborn dump[2] states ua and ub halves the entropy cost.

4.2. Application

Lemma 4.2.1. *Let $L = \{w \in \{a,b\}^* | \ w$ contains "bb" and ends with a "b"$\}$. Then there exists a DFA which recognizes the language $L_2 = \{xw_1 \ldots xw_n | w \in L\}$ with average cost of entropy at most $\frac{1}{2} + \epsilon$ for any given $\epsilon > 0$.*

[1]In the sense that $\chi(ux, \sigma) \leq 1$ for every $\sigma \in \{a,b,x\}$.
[2]In the sense that there is no coming back from them to acceptance.

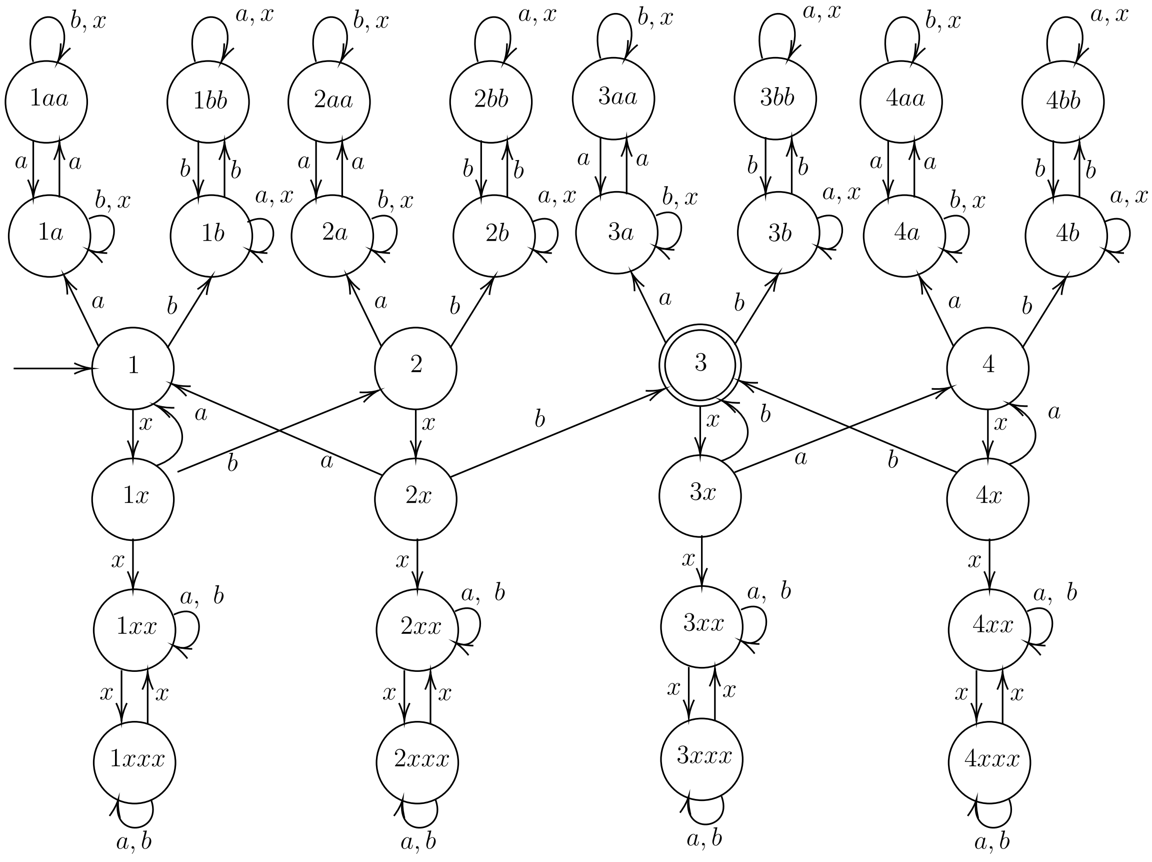

Figure 4.2. The DFA M.

Proof. By [1, Lemma 10], given $\epsilon > 0$, there exists a DFA D_ϵ which recognizes L_2 with an average cost of entropy at most $1 + \epsilon$.

Apply the above construction to D_ϵ: Let $M = (\{a, b, x\}, Q', \delta', 1, \{3\})$ be the DFA where $Q' = Q \cup Qx \cup Qa \cup Qb \cup Qxx \cup Qaa \cup Qbb \cup Qxxx$, and the transition function δ' is as follows:

$$\delta'(q, \sigma) = \begin{cases} qx & \text{if } \sigma = x \\ qa & \text{if } \sigma = a. \\ qb & \text{if } \sigma = b. \end{cases}$$

$$\delta'(qaa, \sigma) = \begin{cases} qaa & \text{if } \sigma = b, x \\ qa & \text{if } \sigma = a. \end{cases}$$

$$\delta'(qx, \sigma) = \begin{cases} \delta(q, \sigma) & \text{if } \sigma = a, b \\ qxx & \text{if } \sigma = x. \end{cases}$$

$$\delta'(qbb, \sigma) = \begin{cases} qbb & \text{if } \sigma = a, x \\ qb & \text{if } \sigma = b. \end{cases}$$

$$\delta'(qa, \sigma) = \begin{cases} qa & \text{if } \sigma = b, x \\ qaa & \text{if } \sigma = a. \end{cases}$$

$$\delta'(qxx, \sigma) = \begin{cases} qxx & \text{if } \sigma = a, b \\ qxxx & \text{if } \sigma = x. \end{cases}$$

$$\delta'(qb, \sigma) = \begin{cases} qb & \text{if } \sigma = a, x \\ qbb & \text{if } \sigma = b. \end{cases}$$

$$\delta'(qxxx, \sigma) = \begin{cases} qxxx & \text{if } \sigma = a, b \\ qxx & \text{if } \sigma = x. \end{cases}$$

Let us call the states in Qa, Qb, Qxx, Qaa, Qbb, and $Qxxx$ the *outer states,* and those in Q and Qx the *inner states* of M. Note that there is no coming back from the outer states to the inner states, hence any computation entering an outer state results in rejection; whereas states in Qx extend the previous transitions in D by appending one x in front of every letter a or b; therefore, the given DFA M_1 recognizes the language L_2.

If $w \in \{a, b, x\}^n$, we want to show that

$$\frac{\sum_{i=1}^{n} \log_2(\chi(u_i, w_i))}{n} \leq \frac{1}{2}.$$

Let j be the smallest index such that u_j is an outer state of M. Thereafter, the only source of entropy is the $\log_2(2) = 1$ entropy caused by the incoming a, b, and x-transitions for states in Qa, Qb, and Qxx, respectively; but, in the worst case, these can be incurred at every 2 steps, leading to the bound

$$\chi(u_j, w_j) + \cdots + \chi(u_n, w_n) \leq \left\lfloor \frac{n - j + 1}{2} \right\rfloor.$$

To compute the rest of the sum, we consider the cases where j is even and odd separately. If j is even, then for this sub-computation to stay in the inner states, the $j-1$-letter long prefix w' of w must be in the form $w' = xw_2 \ldots xw_{j-2}x$, and the corresponding states u_i, $0 \leq i < j$, must belong to Q for i even, and to Qx for i odd. The numbers of incoming transitions for the original states Q did not change; therefore, the entropy cost of running w' on M can be calculated as

$$\begin{aligned}\sum_{i=1}^{j-1} \log_2(\chi_M(u_i, w_i)) &= \log_2(\chi_M(u_1, x)) + \log_2(\chi_M(u_2, w_2)) + \cdots + \log_2(\chi_M(u_{j-1}, x)) \\ &= \log_2(1) + \log_2(\chi_D(u_2, w_2)) + \cdots + \log_2(\chi_D(u_{j-2}, w_{j-2})) + \log_2(1).\end{aligned}$$

Since the entropy cost of the string $w_2 w_4 \ldots w_{j-2}$ of length $(j-2)/2$ can be at most

$(j-2)/2+\epsilon(j-2)/2$ on D, when j is even, cost of w' on M is seen to be bounded by

$$\sum_{i=1}^{j-1} \log_2(\chi_M(u_i, w_i)) \leq \frac{j-2}{2} + \epsilon\frac{j-2}{2}.$$

Similarly, when j is odd, the prefix w' is of the form $w' = xw_2 \dots xw_{j-1}$, and u_i, $0 \leq i < j$, still belongs to Q for i even, and to Qx for i odd, which divides the sum as

$$\begin{aligned}\sum_{i=1}^{j-1} \log_2(\chi_M(u_i, w_i)) &= \log_2(\chi_M(u_1, x)) + \log_2(\chi_M(u_2, w_2)) + \cdots + \log_2(\chi_M(u_{j-1}, w_{j-1})) \\ &= \log_2(1) + \log_2(\chi_D(u_2, w_2)) + \cdots + \log_2(1) + \log_2(\chi_D(u_{j-1}, w_{j-1})).\end{aligned}$$

The entropy cost of $w_2w_4 \dots w_{j-1}$ of length $(j-1)/2$ can be at most $(j-1)/2+\epsilon(j-1)/2$ on D; hence, when j is odd, the entropy cost of w' on M is bounded by

$$\sum_{i=1}^{j-1} \log_2(\chi_M(u_i, w_i)) \leq \frac{j-1}{2} + \epsilon\frac{j-1}{2}.$$

Combining the inner and the outer states, the entropy cost is bounded by

$$\begin{aligned}\sum_{i=1}^{n} \log_2(\chi_M(u_i, w_i)) &= \sum_{i=1}^{j-1} \log_2(\chi_M(u_i, w_i)) + \sum_{i=j}^{n} \log_2(\chi_M(u_i, w_i)) \\ &\leq \max\left\{\frac{j-2}{2}, \frac{j-1}{2}\right\}(1+\epsilon) + \left\lfloor \frac{n-j+1}{2} \right\rfloor \\ &\leq \frac{n}{2}(1+\epsilon).\end{aligned}$$

■

We define $L_k = \{x^{k-1}w_1 \dots x_{k-1}w_n | w_1 \dots w_n \in L\}$ for $k \geq 1$.

Corollary 4.2.2. *For every $k \geq 1$, there is a language, namely L_k, with entropy cost at most $\frac{1}{k} + \epsilon$.*

For the second part of our project, we aim to set a lower bound on the energy complexity of the languages L_k. Namely, we want to prove the following:

Lemma 4.2.3. *Let M be any DFA recognizing the language L_k. Then there exists $\epsilon > 0$ such that for all sufficiently large n, there exists an input string $w = w_1 \dots w_n$ for which*

$$\frac{\sum_{i=1}^{n} \log_2(\chi(q_i, w_i))}{n} \geq 1/k + \epsilon,$$

where $(q_0, q_1, q_2, \dots, q_n)$ is the sequence of states traversed by M during the consumption of w, beginning with the start state q_0.

Proof. Without loss of generality, discard any unreachable state in M. Then, by the Myhill-Nerode Theorem, the states of M fall into $4 + 3(k-1) + 1 = 3k+2$ equivalence classes corresponding to the distinct states of the minimal DFA recognizing L_k, see Figure 4.3 for the case $k = 2$. The general case of $k > 2$ is the straightforward generalization of this DFA with $k-1$ consequent x-transitions needed to reach the original states 1, 2, 3, and 4.

Call a state of M a *trash state* if it is equivalent to the state 0 in the minimal DFA. Inputs of the form $x^{k-1}w_k x^{k-1} w_{2k} \dots x^{k-1} w_{nk}$ where $w_{ik} \in \{a, b\}$, $i = 1, \dots, n$, does not enter trash as at any step, a possibility for being accepted is preserved; whereas any other input enters trash immediately when it breaks this form because every non-trash state has a way of reaching an accepting state, but the input we consider would no longer have that possibility. Hence, M is comprised of one or more trash states connected to a k-partite graph of states. Even more is true: If we call the part containing the initial state the first part, and the part containing $\delta_M(q, x^i)$ the ith part, then a state in the ith part, where $i < k$, has transitions only to another states in the $(i+1)$th part, and those in the kth part has transitions only to the first part. Call the states which are in the first part the *essential states* and those in the other parts the *auxiliary states.*

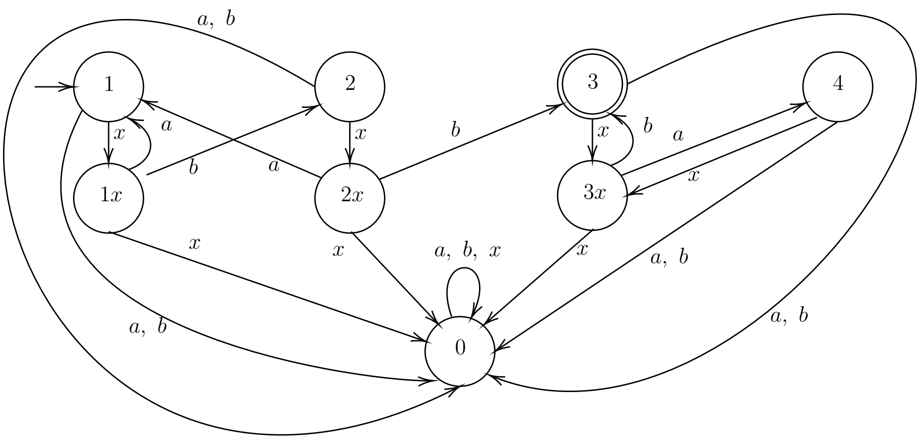

Figure 4.3. Minimal DFA of L_2.

Construct a new DFA M' on top of the essential states of M with the transition function

$$\delta_{M'}(q,a)=\delta_M(q,x^{k-1}a),\ \delta_{M'}(q,b)=\delta_M(q,x^{k-1}b).$$

Since every new transition fits the form $x^{k-1}w_k x^{k-1}w_{2k}\dots x^{k-1}w_{kn}$ above, there is no need for the trash states; and since we have mapped strings of length k to those of length 1, no need for the other parts of the k-partite graph –i.e. the auxiliary states– as well because of the ordered structure of M we alluded to above. In particular, the number of a- and the number of b-transitions entering an essential state are preserved in passing from M to M'. That is to say,

$$\chi_{M'}(q,a)=\chi_M(q,a) \text{ and } \chi_{M'}(q,b)=\chi_M(q,b).$$

The language recognized by this new machine M' is

$$\{w_1\dots w_n\in\{a,b\}^n|x^{k-1}w_1\dots x^{k-1}w_n\in L_k\}=L.$$

By [1, Theorem 11], for all sufficiently large n, there is a string $w=w_1\dots w_n$ such that for some $\epsilon>0$,

$$\log_2(\chi_{M'}(q_1,w_1))+\cdots+\log_2(\chi_{M'}(q_n,w_n))\geq(1+\epsilon)n,$$

where $(q_0,\dots,q_n)$ is the sequence of states M' traverses while consuming w. Then, for auxiliary states p_{ij} where $i=1,\dots,n$ and $j=1,\dots,k-1$, the input $\hat{w}=x^{k-1}w_1\dots x^{k-1}w_n$ makes M traverse the state sequence

$$(q_0,\ p_{1,1},\ \dots,\ p_{1,k-1},\ q_1,\ \dots,\ p_{n,1},\ \dots,\ p_{n,k-1},\ q_n),$$

by the definition of the transition function of M'. The entropy cost of this computation is

$$\begin{aligned}\sum_{i=1}^{n}\left(\sum_{j=1}^{k-1}\log_2(\chi_M(p_{ij},x))+\log_2(\chi_M(q_i,w_i))\right) &\geq \sum_{i=1}^{n}\log_2(\chi_M(q_i,w_i))\\ &= \sum_{i=1}^{n}\log_2(\chi_{M'}(q_i,w_i))\\ &\geq (1+\epsilon)n.\end{aligned}$$

Hence, the average entropy cost of $\hat{w}$ on M is

$$\frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{k-1}\log_2(\chi_M(p_{ij},x))+\log_2(\chi_M(q_i,w_i))\right)}{kn} \geq \frac{(1+\epsilon)n}{kn} = \frac{1}{k}+\frac{\epsilon}{k}.$$

■

5. RESULTS

We have proved the following results:

Lemma 5.0.1. *Let $L = \{w \in \{a,b\}^* |$ w contains "bb" and ends with a "b"$\}$. Then there exists a DFA which recognizes the language $L_2 = \{xw_1 \dots xw_n | w \in L\}$ with average cost of entropy at most $\frac{1}{2} + \epsilon$ for any given $\epsilon > 0$.*

Corollary 5.0.2. *For every $k \geq 1$, there is a language, namely L_k, with entropy cost at most $\frac{1}{k} + \epsilon$.*

Lemma 5.0.3. *Let M be any DFA recognizing the language L_k. Then there exists $\epsilon > 0$ such that for all sufficiently large n, there exists an input string $w = w_1 \dots w_n$ for which*

$$\frac{\sum_{i=1}^{n} \log_2(\chi(q_i, w_i))}{n} \geq 1/k + \epsilon,$$

where $(q_0, q_1, q_2, \dots, q_n)$ is the sequence of states traversed by M during the consumption of w, beginning with the start state q_0.

Theorem 5.0.4. *For every $n > 0$, there exists a language with entropy $1/n$.*

6. CONCLUSION

Our main result is the Entropy Hierarchy Theorem.

Theorem 6.0.1 (Entropy Hierarchy). *For every $n > 0$, there exists a language with entropy $1/n + \epsilon$.*

Thus, we build a complexity hierarchy on the class of regular languages for the entropy cost function.

REFERENCES

1. Kıyak, F. and A. C. C. Say, "Energy Complexity of Regular Languages", , 2023.

2. Landauer, R., "Irreversibility and Heat Generation in the Computing Process", *IBM Journal of Research and Development*, Vol. 5, No. 3, pp. 183–191, 1961.

3. Kondacs, A. and J. Watrous, "On the power of quantum finite state automata", *Proceedings 38th Annual Symposium on Foundations of Computer Science*, pp. 66–75, 1997.

4. Pin, J.-E., "On reversible automata", *LATIN'92: 1st Latin American Symposium on Theoretical Informatics São Paulo, Brazil, April 6–10, 1992 Proceedings*, pp. 401–416, 2005.

5. Li, M. and P. Vitanyi, "Reversibility and Adiabatic Computation: Trading Time and Space for Energy", *Proceedings: Mathematical, Physical and Engineering Sciences*, Vol. 452, No. 1947, pp. 769–789, 1996.

6. Lange, K.-J., P. McKenzie and A. Tapp, "Reversible Space Equals Deterministic Space", *Journal of Computer and System Sciences*, Vol. 60, No. 2, pp. 354–367, 2000.