Random Walks and Universal Sequences

Xiaochen Qi *

February 28, 2013

Abstract

A random walk is a chance process studied in probability, which plays an important role in probability theory and its applications. The first connection linking random walk theory to computer science is the study based on the question of universal sequences, which is related to some problems of computation complexity.

We introduce Stephen Cook's original question on universal sequences, and provide a theorem as the answer to it. Then we demonstrate the proof of this theorem provided by Aleliunas *et al*[1], which depends on the analysis of random walks in undirected graphs.

1 Introduction

The question of universal sequences was proposed by Stephen Cook with the motivation for the proof of lower bounds on the space complexity of the reachability problem.

Cook's question[1] was set on n-vertex d-regular graph G, which means each vertex has a fixed degree d. For each vertex v, let edges incident with v be labelled distinctly with $0, 1, \cdots, d-1$. Note that the labels are assigned arbitrarily, which means that each endpoint of an edge can be labelled in a different manner. A sequence σ in $\{0, 1, \cdots, d-1\}^*$, is said to traverse G from v, if starting from vertex v following sequence σ, one covers all the vertices in G. σ is called $n-universal$ if it can traverse every n-vertex graph with degree d, from every starting vertex v. So the question was: are there always short n-universal sequences? (Here short means of polynomial in n.)

So if short n-universal sequences always exists, then the space complexity of reachability problem in undirected graph is logspace. Let T_n be a two-way finite automaton to solve the readability problem on input n. Just let T_n follow a sequence of move instructions and then

*Dartmouth College, Hanover, NH.

stop, the problem can be solved. Aleliunas *et al*, gave us the result that universal sequences do exists:

Theorem 1.1. *There is an n-universal sequence of length $O(n^3 log\ n)$. (Implied constant depends only on the fixed degree d.)*

To prove this theorem, we need not only to analyze random walk in an undirected graph, but also to use a probabilistic method with a small probability of failure that can be made arbitrarily small. The proof for this theorem is a good example for probabilistic algorithm's application. Before turning to the proof of the existence of universal sequences, we shall look into the properties of random walks on undirected graphs first.

2 Preliminaries

Let G be an n-vertex undirected connect graph. d_i denotes the degree of vertex i. e is the number of edges in G, which is $\frac{1}{2}\sum_i d_i$. Consider a random walk on G, the transition probability P_{ij} is:

$$P_{ij} = \begin{cases} 0 & \text{if } \{i,j\} \text{ is not an edge} \\ \dfrac{1}{d_i} & \text{otherwise} \end{cases}$$

Let $\mathbb{E}_i\tau_j$ denote the expected time starting from i to reach j. $\mathbb{E}_i\tau_i$ is the mean return time of vertex i.

Lemma 2.1. *For each vertex i of G, $\mathbb{E}_i\tau_i = \dfrac{2e}{d_i}$*

Proof. (cf. [2]) Let π_i be the stationary probability of vertex i. Stationary distribution $\pi = (\pi_1, \pi_2, \cdots, \pi_n)$ satisfies:

$$\pi P = \pi \text{ and } \sum_i \pi_i = 1$$

It's easy to verify that $\pi_i = \dfrac{d_i}{2e}$. Note that the stationary probability of a vertex is the reciprocal of its mean return time. Just think one has probability π_i to get back to i, which means it takes $\dfrac{1}{\pi_i}$ time or steps to get back once. □

For any edge $\{i,j\}$ in G, we have $\pi_i p_{ij} = \dfrac{d_i}{2e} \cdot \dfrac{1}{d_i} = \dfrac{1}{2e}$, which shows that the frequency of an edge to be traversed from i to j is $\dfrac{1}{2e}$ in long-run.

Lemma 2.2. *For any adjacent vertices i and j in G, $\mathbb{E}_i\tau_j + \mathbb{E}_j\tau_i \leq 2e$ (commute time from i to j and back to i)*

Proof. Notice that in long-run the mean time it takes to traverse $i \to j$ is $2e$. Actually this is a upper bound for the commute time. Consider the worst case: edge $\{i, j\}$ is a bridge in G, which means before traverse $i \to j$ again, one cannot return to i. Thus the commute time of any two adjacent vertices is no greater than $2e$. □

Define C_G as the expected time to traverse all the vertices in G, also known as the cover time of G.

Lemma 2.3. $C_G \leq 2e(n-1)$

Proof. Let H be a spanning tree of G, then one can traverse all vertices starting and ending at one vertex i. And each edge in H is traversed in each direction once. Let the sequence of vertices traversed be $i_0, i_1, \cdots, i_{2n-2}$. Note that $i = i_0 = i_{2n-2}$. Obviously, the cover time C_G is no greater than the expected time to traverse all the vertices following order in this sequence. The expected traversal time of H is:

$$\mathbb{E}_{i_0}\tau_{i_1} + \mathbb{E}_{i_0}\tau_{i_1} + \cdots + \mathbb{E}_{i_{2n-3}}\tau_{i_{2n-2}} = \sum_{\{i,j\} \in H} \mathbb{E}_i\tau_j + \mathbb{E}_j\tau_i$$

By Lemma 2.2, the sum is no more than $2e(n-1)$ □

Lemma 2.2 and Lemma 2.3 can also be proved by using Chandra's Theorem, that is: the commute time between two vertices s and t (the expected length of a random walk from s to t and back) is precisely characterized by the effctive resistance R between s and t: commute time is $2eR$[3]. For adjacent vertices i, j, the effective resistance is at most 1. It's equal to 1 when the edge $\{i, j\}$ is a bridge. For effective resistance of a graph, from Foster's Theorem[4] we know it's $n-1$. Both are consistent with the proof provide above.

3 The existence of universal sequences

The analysis of random walks on undirected graphs give us a loose upper bound of cover time, which can result in the existence of universal sequences via the probabilistic method.

The main idea is to show that expected number of n-vertex d-regular graphs the $O(n^3 log\ n)$ random sequence fails to traverse is less than 1, which means the universal sequence must exist.

Proof. Let $\tilde{\sigma}$ be a random sequence in $\{0, 1, \cdots, d-1\}^*$ of length:

$$L = 2dn(n-1)(dn+2)\lceil log_2 n \rceil$$

Again, we have to ensure that the each label in this sequence is independent and the d characters are equally likely to be chosen to form the sequence.

We also need random variables to record the success or failure of a walk following $\tilde{\sigma}$ for all the labelled n-vertex d-regular graphs. Let $\tilde{X}_{G,v}$ be the family of random variables for the set of labelled n-vertex d-regular graphs, which is indexed by the specific graph G and some starting vertex v. Define $\tilde{X}_{G,v}$ as follows:

$$\tilde{X}_{G,v} = \begin{cases} 0 & \text{if starting from } v\text{, following } \tilde{\sigma}, \\ & \text{one fails to traverse all the vertices in } G \\ 1 & \text{otherwise, success} \end{cases}$$

Define random variable $\tilde{Y} = \sum \tilde{X}_{G,v}$, the sum of expectation of counter examples of all n-vertex d-regular graph, that is to say, $\tilde{Y}$ is the counter of failures. By definition, if $\tilde{Y} = 0$, then $\tilde{\sigma}$ is n-universal. So the key to prove that n-universal sequences exist, is to show that when $\tilde{\sigma}$ is a random sequence, then $\tilde{Y}$ has to be less than 1.

We have $\mathbb{E}(\tilde{Y}) = \sum \mathbb{E}(\tilde{X}_{G,v})$ because expectation is linear. To prove that $\mathbb{E}(\tilde{Y})$ is less than 1 means to prove that for each G, v, $\mathbb{E}(\tilde{X}_{G,v})$ is less than some value which depends on n and d.

Consider $\tilde{\sigma}$ is concatenated of $N = (dn+2)\lceil log_2 n \rceil$ random subsequences $s_1, s_2, \cdots, s_N$ each with length $l = 2dn(n-1)$. Note that each subsequence is independent with each other, which means there's no overlapping between any of them.

By Lemma 2.3, the cover time of such n-vertex d-regular graph G, $C_G \leq 2e(n-1) = dn(n-1)$. Also by Markov's Theorem[5], $P(\tilde{C}_G \geq t) \leq \dfrac{C_G}{t}$, here $\tilde{C}_G$ is a random variable for the time to traverse all vertices in G, starting from some vertex v. In this case is:

$$P(\tilde{C}_G \geq l) \leq \frac{C_G}{l} \leq \frac{1}{2}$$

This directly indicates that the probability of the the time to traverse all vertices in G with no less than time l is no more than $\frac{1}{2}$, which means if we follow a random variable $\tilde{\sigma}$ with length l, we have the same probability to fail to traverse all the vertices. Hence the expectation of $\tilde{X}_{G,v}$ is no greater than $\frac{1}{2}$.

To compute the probability that none of the subsequences $s_1, s_2, \cdots, s_N$ will cover all the sequences in graph G starting from vertex v, since each subsequence is independent, we have:

$$\mathbb{E}(\tilde{X}_{G,v}) \leq 2^{-N} = 2^{-(dn+2)\lceil log_2 n \rceil} \leq n^{-(dn+2)}$$

We also have a upper bound for the number of labelled n-vertex d-regular graphs. The number of such graphs is less than n^{dn}. Just consider that for each vertex we have less than n^d choices of edges.

Finally, we get:

$$\mathbb{E}(\tilde{Y}) = \sum \mathbb{E}(\tilde{X}_{G,v}) \leq n^{dn} \cdot n \cdot n^{-(dn+2)} = n^{-1} < 1$$

Note that the n in the middle is the number of choices for the starting vertex.

Thus the n-universal sequences of length $O(n^3 log\ n)$ exists. □

4 Conclusions

The proof of existence of n-universal sequences gives an affirmative answer to S. Cook's question and thus proves the lower bound of space complexity of reachability problem in unditected graphs. In comparison, universal sequences for directed graphs are of larger length($O(d^n)$), and the structure of reachability problem for directed graphs is quite different and more complex than that for undirected graphs.

Also there are still some questions open. First is that even we have proved the existence of universal sequences, but no one has ever been able to find one such sequence. That's the magic of probabilistic method, which one can prove the existence without providing one instance. Secondly, whether this bound can be improved is yet to be answered.

References

[1] R. Aleliunas, R.M. Karp, R.J. Lipton, L. Lovász and C. Rackoff. (1979) Random walks, universal traversal sequences, and the complexity of maze traversal, *Proceedings of the 20th IEEE Symposium on Foundations of Computer Science*, pp. 218-233.

[2] F. Göbel and A. Jagers. (1974) Random walks on graphs. *Stochastic Processes and their Applications*, 2:311-336.

[3] Chandra, A. K., Raghavan, P., Ruzzo, W. L, Smolensky, R. and Tiwari, P. (1989) The electrical resistance of a graph captures its commute and cover times. *Proc. of the 21st Annual ACM Symp. on Theory of Computing* 574-586.

[4] Foster, R. M. (1949) The Average impedance of an electrical network. *Contributions to Applied Mechanics* (Reissner Anniversary Volume), Edwards Bros., Ann Arbor, Mich. 333-340.

[5] Markov, A. (1884) On certain applications of algebraic continued fractions, Ph.D. thesis, St Petersburg.