LEARNING MODELS FOR LANGUAGE ACQUISITION

SHASHI MITTAL, HARISH KARNICK

Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur, India 208016.

mshashi@iitk.ac.in, hk@cse.iitk.ac.in

In this paper, we present a model of language acquisition which can be used to explain how children learn a grammar by interacting with their surroundings. We build upon the model proposed by Komarova et al in the context of evolution of grammars. We test our model for two situations : One, in which an individual is trying to learn a grammar in an environment where everybody uses the same grammar, and the other in which different groups in the population use different grammars.

1. Introduction

Komarova et al (Komarova, Niyogi, & Nowak, 2001; Komarova & Nowak, 2002) have proposed a computational model for the evolutionary dynamics of grammar acquisition in a population of individuals. Along similar lines we propose a model for grammar acquisition for a new language learner. The fact that a human learner acquires a language and in particular its grammar, coupled with impossibility results from computational learning theory (Gold, 1967; Angluin & Kharitonov, 1995) imply that there exists an inherent learning bias that makes learning the grammar of a language from limited input computationally feasible. One possibility is the presence of a finite set of candidate grammars one of which gets selected as the most appropriate based on linguistic input (Nowak, Komarova, & Niyogi, 2002). Here, we study language acquisition mainly in the context of learning the correct grammar from the available linguistic data.

2. The grammar system

Komarova et al's model is briefly as follows:

- **UG** = Finite search space of candidate grammars, G_1, G_2, $\ldots G_n$, also called the *Universal Grammar*.
- s_{ij} = Probability that a speaker who uses G_i utters a sentence compatible with G_j. It represents the similarity between different grammars.
- $F(G_i, G_j)$ = $(s_{ij} + s_{ji})/2$, the payoff for mutual understanding.

- x_i = Relative abundance of individuals who use grammar G_i.
- $f_i = \sum_{j=1}^{n} x_j F(G_i, G_j)$, This is the payoff for an individual using grammar G_i.
- $\{q_{ij}\}$ is the stochastic learning matrix, q_{ij} denotes the probability that a child born to an individual using G_i will develop G_j.

The population dynamics of grammar acquisition (Komarova et al., 2001; Komarova & Nowak, 2002) is:

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} f_i q_{ij} x_i - \phi x_j, j = 1, 2, \ldots, n. \tag{1}$$

where $\phi = \sum_{i=1}^{n} f_i x_i$ denotes the average fitness, or *grammatical coherence* of the population. In general, ϕ is a number between 0 and 1.

To model language acquisition dynamics for a single individual, we introduce the following concepts:

- A = number of individuals in the population
- p_{ij} = probability that the ith individual uses G_j at a particular instant in time. Clearly, $\sum_{j=1}^{n} p_{ij} = 1$, and also $x_j = \frac{1}{A} \sum_{k=1}^{A} p_{kj}$.
- $\psi_i = \sum_{j=1}^{n} p_{ij} f_j$, this represents the payoff for the ith individual.

3. Grammar acquisition dynamics of an individual

Following Komarova et al, the learning model for the evolutionary dynamics of grammar acquisition is: Let $\{q_{ij}\}$ be the stochastic learning matrix, where q_{ij} denotes the probability that an individual using grammar G_i will switch to using grammar G_j in the next turn. Note that this interpretation of the stochastic learning matrix is different from the one described in the previous section. Using this stochasticity matrix, the learning dynamics of an individual is given by:

$$\frac{dp_{ij}}{dt} = \sum_{k=1}^{n} f_k q_{kj} p_{ik} - \psi_i p_{ij}, \quad i = 1, \ldots, A, \; j = 1, \ldots, n \tag{2}$$

In all, we have $A \times n$ differential equations, where p_{ij} corresponds to the probability that individual i uses grammar G_j from the **UG**.

4. A simple learning model

The simplest learning model is to assume that all q_{ij} are constants. To simplify our analysis, we assume the q matrix is symmetric, and is given by

$$q_{ii} = q, \ , i = 1, \ldots, n \tag{3}$$

$$q_{ij} = \frac{1-q}{n-1}, i \neq j \tag{4}$$

Further, we assume a *fully symmetrical system*, that is

$$s_{ij} = s, i \neq j, 0 < s < 1 \tag{5}$$

4.1. *Language acquisition dynamics when all population members use the same grammar*

This problem can be formulated as follows : We assume that there are n grammars in the universal grammar set, and there are A individuals in the population. Without loss of generality, we assume that individuals from 1 to $A-1$ have chosen grammar G_1, and we are interested in studying the dynamics of the Ath individual. Assuming that the Ath individual uses all the grammars except G_1 with uniform probability, the above equation reduces to the following form

$$\frac{dp_{A1}}{dt} \ = \ \alpha p_{A1}^3 + \beta p_{A1}^2 + \gamma p_{A1} + \delta \tag{6}$$

where, $\alpha = -(\frac{1-s}{A})$, $\beta = (1-s)(\frac{q+1-A}{A})$, $\gamma = q(1-\frac{1-s}{A}) - \frac{s(1-q)}{n-1} - s$ and $\delta = \frac{s(1-q)}{n-1}$.

The initial condition for the equation is

$$p_{A1} = 1/n, \ t = 0 \tag{7}$$

that is, initially each grammar has equal probability of being used. We are interested in studying the behavior of this initial value problem, in particular we want to see whether p_{A1} always attains an equilibrium, if so what is the equilibrium value and how do the parameters s, q, A and n influence the acquisition process. Mathematica was used to study the behaviour of the differential equations.

The probability p_{A1} (henceforth referred to as p in this section for convenience) always converges, though the value to which it converges depends upon the values of the parameters. The value of p reaches 1.0 only if $q = 1.0$ (i.e. learning fidelity is perfect). The effect of changing the parameters q, s, n and A can be summarized as follows :

- If the value of q is increased, keeping other variables fixed, the value of p converges to a higher value as shown in Fig 1. At $q = 1.0$, the final value of p is 1.0, irrespective of the value of s.
- If only s is changed, the value to which p converges decreases, and so does the rate of convergence, as shown in Fig. 2. ($q = 0.7$, $n = 10$ and $A = 10$).
- With increasing n, the convergence is attained at a slower rate, although it always converges to the same value.
- Changing the value of A does not show any significant impact on grammar acquisition dynamics.

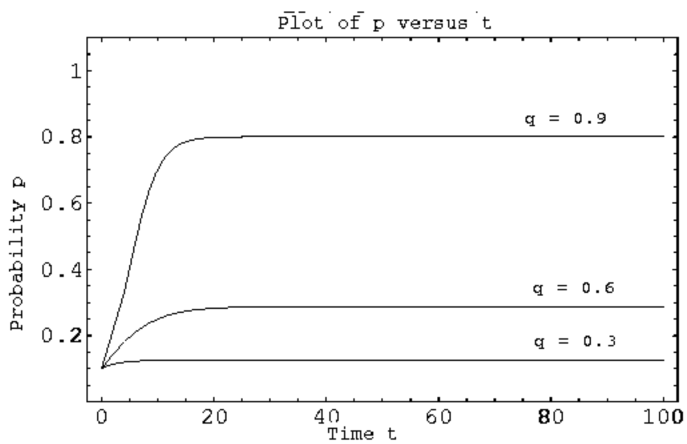

Figure 1. Plot of p versus t when q is varied.

4.2. *Learning mechanisms*

A learning mechanism defines the dependence of $\{q_{ij}\}$ on N, the number of learning events. The results for two learning algorithms, both of which have been extensively studied in the literature, are presented here.

- **Memoryless learning:** The learner starts with a randomly chosen hypothesis (say G_i) and stays with this hypothesis as long as the sentence heard is compatible with this hypothesis. If a sentence is not compatible, the learner randomly chooses another grammar from the **UG**. The process stops after N sentences. For a fully symmetrical system, the dependence of q on N is

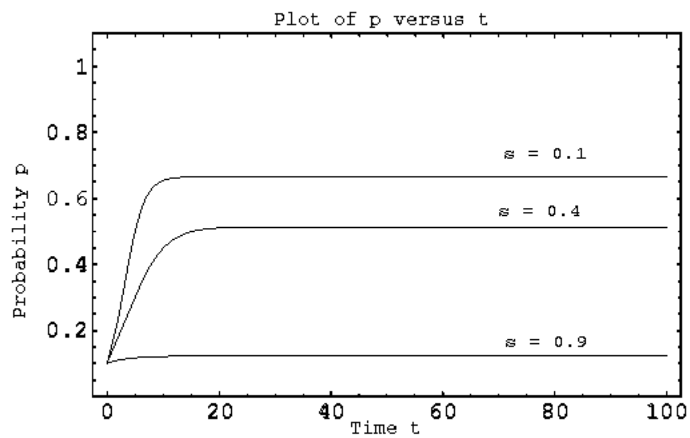

Figure 2. Plot of p versus t when s is varied

given by

$$q = q_{ii} = 1 - (1 - \frac{1-s}{n-1})^N \ \frac{n-1}{n} \tag{8}$$

- **Batch learning:** The Batch learner is first exposed to and memorizes all N sentences and then chooses a grammar from the **UG** that is most compatible with the input. For a fully symmetrical system,

$$q = q_{ii} = \frac{(1-(1-s^N)^n)}{s^N n} \tag{9}$$

It can be seen that the value of q will be higher for the batch learner compared to the memoryless learner, for the same values of s, n and A. Human learning is likely to be intermediate between these two algorithms, and therefore human performance is expected to lie somewhere between these two. Fig. 3 shows the plot of p_{A1} when $s = 0.4$, $N = 40$, $n = 25$ and $A = 10$. The memoryless algorithm converges to $p = 0.40$, whereas the batch learner algorithm converges to $p = 0.85$. The corresponding values of q for the two algorithms are 0.65 and 0.92 respectively.

4.3. *Language acquisition dynamics when different population members use different grammars*

We formulate this problem as follows : The **UG** has n grammars in . Members 1,2,..., A use grammar G_1, members $A+1$, $A+2$, ..., $2A$ use G_2, and so on.

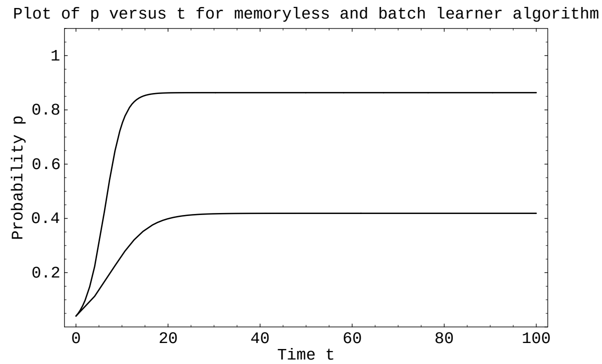

Figure 3. Plots for memoryless and batch learner algorithm

The $(nA + 1)$th member is the learner and is interacting uniformly with all the groups. The dynamics of grammar acquisition for this member is given (in a fully symmetrical situation) by:

$$\frac{dp_j}{dt} = (\frac{1-q}{n-1}) \sum_{k=1,k\neq j}^{n} f_k p_k + q f_j p_j - (\sum_{k=1}^{n} f_k p_k) p_j$$

we use p_i for the probabilities of the learning individual.

$$x_j = \frac{\sum_{k=1}^{nA+1} p_{kj}}{nA+1} = \frac{A + p_j}{nA+1}$$

$$f_i = \frac{[(n-1)s + 1]A + s(1-p_i) + p_i}{nA+1}$$

$$\psi = \sum_{k=1}^{n} f_k p_k$$

For this situation, irrespective of the initial values of the probabilities, and the values of s, q, n and A, it is observed that the probability values $p_1, \ldots, p_n$ all converge to the same value. Figure 4 shows the plot for one such case, when $s = 0.2$, $n = 3$, $A = 5$ and $q = 0.8$.

5. A simulated annealing learning model

In the learning models described above, the value of q (the learning coefficient) had been kept constant. In the simulated annealing learning model, the value of q

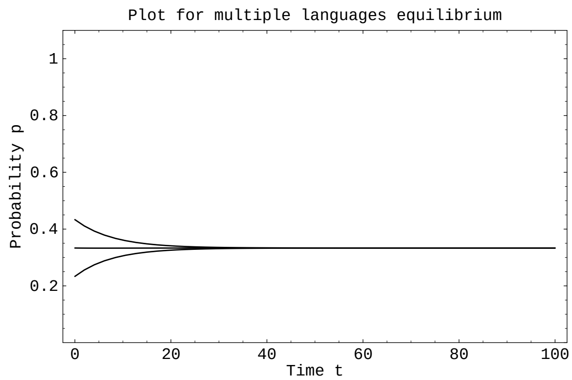

Figure 4. Plots for multiple languages case

changes with time and is given by:

$$q = e^{(\psi-1)/kt} \qquad (10)$$

where $\psi = \sum_{k=1}^{n} f_k p_{ik}$, and k is a constant (fixed at 1.0).
Such a choice of q satisfies the following two important properties (note that q is q_{ii}).

1. If the individual's grammatical coherence is high (i.e. ψ is close to 1, then q is close to 1, i.e. the individual has a lower tendency to switch to another grammar.

2. As time progresses, q tends to 1, i.e. if learning has taken place initially, then there is less likelihood the individual will change to another grammar. However, when t is small q is close to 0 and the learner is likely to switch grammars during early learning.

For the case when all the population members use the same grammar and the simulated annealing learning model is used, the probability of using that particular grammar *always converges to 1*, irrespective of the values of s or n. For the multilingual environment case, the probabilities tend to converge to the same values initially, but subsequently only one of the grammars attains the probability 1 and for other grammars the probability of usage tends to zero. This is shown in Fig. 5.

6. Conclusion

In this paper, we have presented two possible models of learning (the simple learning algorithm and the simulated annealing learning algorithm), and analyzed the behavior of a learner for monolingual and multilingual environments. The simulated annealing model has the interesting consequence that the learner learns a

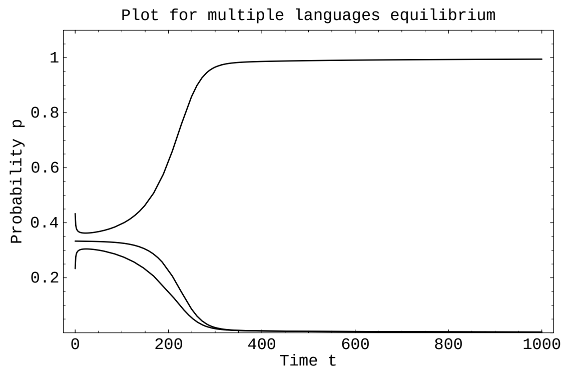

Figure 5. Plot for the multiple language case when simulated annealing model is used.

single language perfectly. The work can be extended by studying the dynamics using more realistic assumptions for the variation of q with time and for the nature of interaction between the learner and the group of mature individuals.

References

Angluin, D., & Kharitonov, M. (1995). When won't membership queries help? *J. Comput. Syst. Sci.*, *50*(2), 336-355.

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, *10*(5), 447-474.

Jain, S., Osherson, D., Royer, J. S., & Sharma, A. (1999). *Systems that learn* (2nd ed.). Cambridge, MA, USA: The MIT Press.

Komarova, N. L., Niyogi, P., & Nowak, M. A. (2001). The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology*, *209*, 43-59.

Komarova, N. L., & Nowak, M. A. (2002). Population dynamics of grammar acquisition. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language.* New York, NY, USA: Springer-Verlag New York.

Mittal, S. (2005). *Investigating learning models for language acquisition* (Tech. Rep.). Indian Institute of Technology, Kanpur. Available at http://www.cse.iitk.ac.in/reports/view.jsp?colname=446.

Niyogi, P. (2004). *The computational nature of language learning and evolution.* In press. Available at http://people.cs.uchicago.edu/ niyogi/Book.html.

Nowak, M., Komarova, N., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature*, *417*(6), 611-617.