Journal of Mathematical Psychology 76 (2017) 1–6

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Adding population structure to models of language evolution by iterated learning

Andrew Whalen [a,*], Thomas L. Griffiths [b]

[a] University of Edinburgh, UK

[b] University of California, Berkeley, United States

HIGHLIGHTS

- Bayesian Iterated Learning converges to the prior in structured populations.
- We characterize the rate at which populations approach the stationary distribution.
- Population structure increases the probability that neighbors share a language.

ARTICLE INFO

Article history:
Received 20 June 2016
Received in revised form
16 October 2016

Keywords:
Bayesian models
Voter models
Iterated learning
Language evolution
Cultural evolution

ABSTRACT

Previous work on iterated learning, a standard language learning paradigm where a sequence of learners learns a language from a previous learner, has found that if learners use a form of Bayesian inference, then the distribution of languages in a population will come to reflect the prior distribution assumed by the learners (Griffiths and Kalish 2007). We expand these results to allow for more complex population structures, and demonstrate that for learners on undirected graphs the distribution of languages will also reflect the prior distribution. We then use techniques borrowed from statistical physics to obtain deeper insight into language evolution, finding that although population structure will not influence the probability that an individual speaks a given language, it will influence how likely neighbors are to speak the same language. These analyses lift a restrictive assumption of iterated learning, and suggest that experimental and mathematical findings using iterated learning may apply to a wider range of settings.
© 2016 Elsevier Inc. All rights reserved.

Language changes; English today is slightly different from a hundred years ago, and radically different from a thousand years ago. An important cause of language change is the variation that occurs during the language learning process (see, e.g., DeGraff, 2001). One of the major tools that has been used to study the impact of language learning on the structure of languages is the iterated learning model (Kirby, 2001). In iterated learning, a set of simulated learners each learn language from the utterances of other learners and then produce utterances themselves that are provided to other learners. Repeating this process, the learners reshape the language. Simple learning algorithms can lead to significant changes, increasing the regularity of languages (Brighton, 2002; Kirby, 2001; Smith, Kirby, & Brighton, 2003) and expressing or even emphasizing the biases of learners (Griffiths & Kalish, 2007; Kirby, Dowman, & Griffiths, 2007).

The simplest iterated learning model – the case that submits most easily to mathematical analysis – is the transmission chain, in which each learner learns from the previous learner and generates utterances for the next. However, more complex models are possible. Exploring these models is important in two ways. First, it lets us establish the generality of results obtained for transmission chains, which represent the majority of previous analyses. Second, it allows us to explore phenomena that only emerge in more complex models. For example, speakers of the same language tend to cluster together spatially — something that is hard to explain using transmission chains.

In this paper, we explore how more complex population structures influence the outcome of iterated learning. We begin by introducing a formal framework for analyzing iterated learning in which learning is modeled as Bayesian inference. We then build on previous analyses of transmission chains by Griffiths and Kalish (2007), showing that similar analytic results can be obtained with populations where the relationships between learners can

* Correspondence to: The Roslin Institute, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK.
E-mail address: awhalen@roslin.ed.ac.uk (A. Whalen).

http://dx.doi.org/10.1016/j.jmp.2016.10.008
0022-2496/© 2016 Elsevier Inc. All rights reserved.

be expressed as a heterogeneous graph. We verify these results using simulations with two-dimensional lattices, small-world graphs (Watts & Strogatz, 1998) and scale-free graphs (Barabasi & Albert, 1999), population structures that mimic some of the properties of real populations. These simulations show that neighbors in a graph are more likely to share the same language than is expected by chance. To quantify this effect we utilize techniques developed for voter models (Castellano, 2012; Sood, Antal, & Redner, 2008) and show that although the graphical structure of a population does not change how likely an individual learner speaks a certain language, it does impact how likely it is that neighbors will be able to communicate.

1. Iterated Bayesian learning

In the simplest iterated learning model, a population is assumed to be a series of parallel transmission chains. At each step in the chain, a learner learns a language from a single teacher and then transmits a language to a single student. The dynamics of this process depend on the learning algorithm that is used by the students.

One way to specify a learning algorithm is to assume that learners use a form of Bayesian inference (Griffiths & Kalish, 2007). Adopting a language then becomes a statistical inference task where the inductive biases of learners – those factors other than the data that lead them to favor one language over another – are expressed as a prior probability distribution over languages. Under this assumption, learners choose to speak a language, L, based on hearing linguistic data, D. We assume that the probability of speaking L is the same as the posterior probability of the language, calculated using Bayes' rule,

$$p(L|D) = \frac{p(D|L)p(L)}{p(D)}, \tag{1}$$

where $p(L)$ is the prior probability of the language, which may not be equal across languages.

Griffiths and Kalish (2007) showed that for transmission chains the probability that a learner speaks a language, L, after a large number of generations is the same as the prior probability of the language, $p(L)$. Formally, the stationary distribution of the resulting stochastic process is the prior distribution over languages. This result is interesting because it suggests that the variation observed in modern languages can be directly connected to the inductive biases of human language learners. Kirby et al. (2007) expanded on this result, showing that variations on Bayesian learning in which learners are more likely to choose languages with higher posterior probabilities can exaggerate the impact of the prior on the stationary distribution, allowing weak inductive biases to have a strong effect on the structure of the languages produced by iterated learning.

However, this simplest iterated learning model may not accurately represent real populations. To explore the generality of these results, Smith (2009) relaxed the assumption of learning from a single teacher and examined populations of learners who learned a single language from multiple teachers. Using simulations, Smith showed that the language such learners acquire is highly dependent on the initial distribution of languages in a population, and more weakly influenced by prior probabilities. Burkett and Griffiths (2010) pursued these results further, and found that if learners could learn multiple languages from multiple teachers, the distribution of languages in the population over a number of generations will still mirror the prior probability of each language. Convergence to a stable equilibrium that is not the prior distribution can also occur if fitness is added into the model (Kalish, 2007).

In the remainder of the paper, we relax a different assumption and consider learners in a structured population who each learn from a single teacher. The goal of this model is to examine whether the structure of a population will affect the long-term distribution of languages in the population.

2. Introducing population structure

A natural way to capture population structure in cultural evolution is to analyze evolutionary dynamics on graphs, where each node is an agent and edges indicate connections between those agents (e.g., Nowak, 2006). In this section, we analyze iterated Bayesian learning on heterogeneous graphs.

2.1. Bayesian language learning on graphs

Represent a population as a set of N learners arranged on a graph. Each learner speaks one of two languages, L_0 or L_1. Population dynamics are included using a birth–death process: at each time step, a random learner is replaced by a novice learner, the novice learner randomly selects a neighbor, hears an utterance from them, and selects a language based on that utterance. This birth–death process is an abstraction of the biological and cultural processes that shape when and how a learner learns a new language. Although a "birth" may represent an actual birth of a new learner, it might also represent an individual who has chosen to change the language they speak.

Under a Bayesian learning algorithm, learners adopt a language based on a linguistic utterance, D, by selecting a language proportional to the posterior probability of each language,

$$p(L_i|D) = \frac{p(D|L_i)p(L_i)}{p(D|L_0)p(L_0) + p(D|L_1)p(L_1)}. \tag{2}$$

We assume that each utterance is consistent with either L_0 or L_1, and when asked to speak, a teacher correctly produces an utterance consistent with their language with probability $1 - \epsilon$, where ϵ represents an error rate in production. If an utterance, D, is consistent with a language, L_i, then $p(d|L_i) = 1 - \epsilon$. Innate linguistic preferences are included through the prior probability of each language, $p(L_i)$.

2.2. Stationary distribution of languages

In this section, we demonstrate that when learning from a single teacher on heterogeneous graphs, the probability that a specific learner speaks a language after many generations is the same as the prior probability of that language. This extends the result that Griffiths and Kalish (2007) proved for transmission chains to more complex population structures.

An intuition for this result can be obtained by re-imagining the transmission of languages across a graph as a set of chains. In each update, we consider updating the value of a single learner by having that learner learn from a teacher. If we look back in time, that teacher learned their language from someone else, so consider the teacher's teacher. We can then construct a chain of teacher–learner pairs from any individual back to one of the individuals in the initial population. This chain is akin to a transmission chain. The probability that the learner at the end of a chain speaks a language should thus converge to the prior distribution as the chain gets longer.

To make this intuition more precise, we introduce the notion of a Markov process: a process where the probability of future states depends only on the current state. The birth–death process we describe above is a Markov process: each update only depends on the current languages that the learners have adopted, not on the languages spoken by deceased learners. This process is also ergodic: because of the noise in transmission, each learner has a

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات