



UNIVERSIDAD DE BUENOS AIRES  
Facultad de Ciencias Exactas y Naturales  
Departamento de Computación

# Análisis bayesiano del aprendizaje en comunidades de video juegos

Tesis presentada para optar al título de Doctor de la  
Universidad de Buenos Aires en el área Ciencias de la Computación

**Lic. Gustavo Andrés Landfried**

Director de tesis: Dr. Esteban Mocskos  
Codirector de tesis: Dr. Diego Fernández Slezak  
Consejero de estudios: Dr. Hernán Melgratti

Lugar de trabajo: Departamento de Computación, Facultad de Ciencias Exactas y Naturales

Buenos Aires, Abril de 2022

Fecha de defensa:

Firma

# Análisis bayesiano del aprendizaje en comunidades de video juegos

## Resumen

Esta es una tesis en ciencias de la computación motivada por una pregunta antropológica. Entender las propiedades de los sistemas de información cultural es uno de los problemas fundamentales de la antropología, relevante para las ciencias de la computación y la inteligencia artificial multiagente: la sociedad puede ser vista como un sistema de procesamiento distribuido de información, y la cultura como el emergente de su intercambio de información. Hasta ahora, las respuestas con mayor influencia para la antropología se han propuesto a través de modelos matemáticos simples. De allí surgieron, por ejemplo, las hipótesis de que la capacidad de acumulación cultural es proporcional al tamaño efectivo de la población, y que estructuras con moderada conexión favorecen la innovación cultural. Si bien los resultados de estos modelos sirven para ganar intuición de procesos complejos, sus resultados son analogías que no están destinadas a realizar predicciones y por lo tanto poseen una débil corroboración empírica.

En la última década, las ciencias de la computación han tenido un impacto profundo en casi todas las disciplinas científicas, produciendo la emergencia de la ciencia de datos. Esta tesis se enmarca en esta nueva forma de interdisciplina, que pone a prueba los métodos computacionales en base al desempeño que tienen sobre datos reales. Para ello, decidimos centrarnos en las comunidades de videojuegos en línea debido a que ellas son un lugar privilegiado para estudiar cómo cambian las estrategias en el tiempo. En el transcurso de la tesis, esta pregunta fue mutando hasta enfocarse en el estudio de los factores sociales que afectan el aprendizaje individual. En este contexto nos propusimos responder las siguientes preguntas. ¿Cuál es la mejor forma de medir el aprendizaje de un individuo en el tiempo? ¿Cuál es la relación entre la formación de equipos y el aprendizaje individual a largo plazo? ¿Cuál es el efecto que la posición topológica de un individuo en la red de intercambios dinámica tiene sobre el aprendizaje individual?

Si bien el aprendizaje automático funciona muy bien en ciertos contextos, puede colapsar si los datos se desvían un poco de lo que el modelo está acostumbrado. Las ciencias, sin embargo, buscan verdades con validez universal. Para no mentir en contextos de incertidumbre debemos no afirmar más de lo que sabemos, sin decir menos. Por ello, la aplicación estricta de las reglas de la probabilidad (enfoque bayesiano) es el método con el que se fundamentan las verdades en contextos de incertidumbre. Al evaluar hipótesis mutuamente contradictorias (A y no A) la sorpresa, única fuente de información, actúa como el único filtro de las creencias previas, permitiendo no decir más de lo que se sabe incorporando toda la información disponible (maximiza incertidumbre compatible con los datos). Si bien hubieron algunas partes de la tesis que no pudimos evaluar todo el espacio de hipótesis, hicimos un esfuerzo por ajustarnos a este ideal.

En el primer trabajo usamos el modelo bayesiano de habilidad más utilizado en la industria del video juego para estudiar una comunidad en el que las personas podían jugar individualmente o en equipos. Allí encontramos que jugar en equipo está asociado a mayor aprendizaje a largo plazo, y que mantener un equipo estable está asociado a mayor velocidad de aprendizaje. En el segundo trabajo, replicamos el estimador de habilidad y descubrimos que ninguno de los modelos disponibles en aquel momento podía obtener estimaciones iniciales fiables ni garantizaban comparabilidad en el tiempo debido a una incorrecta propagación de la información histórica, solo en la dirección del pasado hacia el futuro. En este contexto, implementamos un modelo (disponible hoy en Julia, Python y R) que al propagar correctamente la información histórica es capaz de proporcionar estimaciones con baja incertidumbre en todo momento, asegurando la comparabilidad histórica. En un tercer trabajo, estudiamos la evolución de una red de partidas en el juego de Go durante un periodo de ocho años y encontramos, con el nuevo estimador, que la posición de los individuos en la red tiene un efecto de segundo orden sobre el aprendizaje en los personas que están en el medio del proceso de aprendizaje, ausente entre novatas y expertas.

**Palabras claves:** Ciencias Sociales Computacionales, Inferencia bayesiana, Cultura, Habilidad, Aprendizaje, Comunidades virtuales, Videojuegos

# Bayesian analysis of learning in video game communities

## Abstract

This thesis in computer science is motivated by an anthropological question. One of the fundamental problems of anthropology, relevant to computer science and multi-agent artificial intelligence, is to understand the properties of cultural information systems. Society can be viewed as a distributed information processing system, and culture as the emergent result of information exchange. Until now, the most influential answers to these questions in anthropology have been proposed using simple mathematical models. They have led, for example, to the hypotheses that the capacity for cultural accumulation is proportional to the effective population size, and that moderately connected structures favor cultural innovation. While the results of these models serve to gain intuition of complex processes, their results are analogies not intended to make predictions and therefore have weak empirical support.

In the last decade, computer science has had a profound impact on almost all scientific disciplines, producing the emergence of data science. This thesis is part of this new form of interdisciplinary, which evaluates computational methods based on their performance on real data. To this end, we decided to focus on online gaming communities because they are a valuable source of information to study how strategies change over time. In the course of the thesis, this question gradually evolved into the study of the social factors that affect individual learning. In this context, we set out to answer the following questions: What is the best way to measure individual's learning curves over time? What is the relationship between team formation and long-term individual learning? What is the effect that an individual's topological position in the dynamic exchange network has on individual learning?

While machine learning works very well in certain contexts, it can crash if the data deviates a bit from what the model is accustomed to. The sciences, however, seek truths with universal validity. To avoid lying in contexts of uncertainty, we must not assert more than we know, without saying less. Therefore, the strict application of the rules of probability (Bayesian approach) is the method by which truths are grounded in contexts of uncertainty. When evaluating mutually contradictory hypotheses (A and not A) surprise, the only source of information, acts as the filter of prior beliefs, allowing not to say more than what is known by incorporating all available information (maximizes uncertainty compatible with the data). Although there were some parts of the thesis that we could not evaluate the entire hypothesis space, we made an effort to adhere to this ideal.

In the first paper we used the Bayesian skill model most commonly used in the gaming industry to study a community in which people could play individually or in teams. There we find that playing in teams is associated with higher long-term learning, and that maintaining a stable teammate is associated with faster learning. In the second paper, we replicated the skill estimator and discovered that none of the models available at that time could obtain reliable initial estimates nor did they guarantee comparability over time due to an incorrect propagation of historical information, only in the direction from the past to the future. In this context, we implemented a model (now public in Julia, Python and R) that, by correctly propagating historical information, is able to provide estimates with low uncertainty at all times, ensuring historical comparability. In a third paper, we studied the evolution of a network of games in the game of Go over a period of eight years and we found, thanks to the new estimator, that the position of individuals in the network has a second-order effect on learning for individuals who are in the middle of the learning process, absent among novices and experts.

**Keywords:** Computational social science, Bayesian inference, Culture, Skill, Learning, Virtual communities, Video games