

Objetivos.

El objetivo del TP es desarrollar un algoritmo de repetición espaciada (SRS) que, en conjunto con un método de inteligencia computacional basado en algoritmos evolutivos que encuentre los mejores parámetros del mismo, presente un buen rendimiento frente a un dataset provisto por Duolingo para el desarrollo e investigación de esta clase de algoritmos. Las medidas de calidad empíricas para verificar la eficiencia de algoritmos para un dataset de este tipo son descritas en Tabibian y otros [1].

El algoritmo a utilizar empleará modelos teóricos de la memoria (en el área de la psicología [2][3]) y otros enfoques más prácticos y de naturaleza heurística que explotan la disponibilidad de datos adquiridos en programas que quieren aprovechar estos fenómenos psicológicos.

Motivación y trabajos anteriores.

Para poder aprender y retener en la memoria distintos conocimientos, es necesario realizar varias actividades de repaso hasta que este se encuentre lo suficientemente “fijado” y disponible para acceder o recuperar en algún momento posterior.

Estas actividades de repaso conllevan tiempo y esfuerzo y, naturalmente, resulta de gran interés determinar cómo influyen en la retención del conocimiento una vez aprendido. Existen numerosos estudios en el área, el primero de ellos del cual se tiene conocimiento es el de Ebbinghaus, que mediante simples experimentos de memorización-repaso-evaluación, comprobó que la capacidad de recordar algún ítem en particular disminuye de forma rápida al principio y luego más lentamente.

Este fenómeno luego fue caracterizado mucho mejor en experimentos controlados, como los descritos en el trabajo de Cepeda y otros [2], y desde entonces surgió el problema de cómo utilizar este modelo de funcionamiento de la memoria para mejorar la retención de estos *componentes de conocimiento* o KC, como son caracterizados en el trabajo de Lindsey y otros [4].

Los SRS existen desde hace muchos años, con sus antecedentes analógicos (las *flashcards* o el sistema de Leitner) como inspiración. Éstos se basan en el *spacing effect*, que bien descrito se encuentra ya por la psicología, como se explicó anteriormente. Pero estos SRS suelen utilizar algoritmos heurísticos simples, con pocas garantías con respecto a su eficiencia, motivo por el que se han hecho investigaciones como las de Tabibian y otros [1] para encontrar un algoritmo de SRS que sea óptimo *dado un modelo de memoria cualquiera*.

Nosotros proponemos con este trabajo un algoritmo de SRS heurístico, basándose en un modelo teórico-práctico de la memoria descrito por Lindsey y otros [4], pero optimizado mediante un algoritmo evolutivo que utiliza las medidas de calidad empíricas que se pueden obtener de los datos provistos libremente por el software Duolingo.

Propuesta de solución.

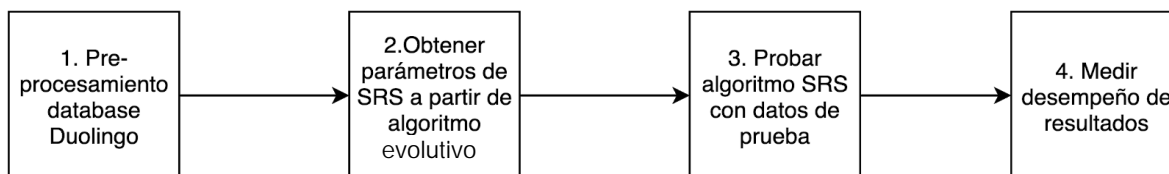


Figura 1 - Diagrama de bloques

En primer lugar, es necesario definir de manera empírica, qué tan eficiente es un algoritmo (a partir de los datos provisto por Duolingo) en definir un schedule de estudio mediante una *función de eficiencia*.

Para esto, tendremos en cuenta la similitud entre el schedule seguido por el usuario (schedule real) y el schedule provisto por nuestro algoritmo (schedule teórico) y, a su vez, una medida del rendimiento en almacenar la información en su memoria siguiendo dicho schedule. Lo ideal buscado en esta comparación es que, a mayor similitud entre los schedules, mayor sea el rendimiento de la persona.

Algo a tener en cuenta es que, en la base de datos provista por Duolingo, pueden existir casos de personas que realizan muy pocas revisiones de los ítems. Para esto, limitaremos los datos a solo aquellos usuarios que hayan realizado al menos 30 revisiones, y para aquellos ítems que hayan sido revisados al menos 30 veces. (Figura 1 - 1)

En segundo lugar, se diseñará un algoritmo del tipo SRS, genérico, que permita modelar el espaciado entre las revisiones de cada ítem. Este algoritmo, contendrá una serie de parámetros a encontrar mediante el algoritmo evolutivo, y otros, que se irán adaptando a la persona en particular que esté siendo analizada, de forma adaptativa (Por ejemplo, la capacidad de aprendizaje de la persona). Los parámetros que corresponden al algoritmo evolutivo, quedarán fijos para todas las personas, y los del SRS cambiarán en base a cada una de ellas. (Figura 1 - 2)

En tercer lugar, se utilizarán los parámetros obtenidos en el paso anterior para ejecutar el algoritmo SRS, obteniendo así, el espaciado esperado entre las revisiones de los ítems por persona.

Finalmente, se comparan los resultados obtenidos, con los almacenados en el dataset de Duolingo, utilizando la misma *función de eficiencia* referida anteriormente.

Referencias bibliográficas

- [1] Tabibian, Behzad, Utkarsh Upadhyay, Abir De, Ali Zarezade, Bernhard Schölkopf, y Manuel Gomez-Rodriguez. "Enhancing human learning via spaced repetition optimization". *Proceedings of the National Academy of Sciences of the United States of America* 116, núm. 10 (el 5 de marzo de 2019): 3988–93.
<https://doi.org/10.1073/pnas.1815156116>.
- [2] Cepeda, Nicholas J., Edward Vul, Doug Rohrer, John T. Wixted, y Harold Pashler. «Spacing Effects in Learning: A Temporal Ridgeline of Optimal Retention». *Psychological Science* 19, n.º 11 (noviembre de 2008): 1095-1102.
<https://doi.org/10.1111/j.1467-9280.2008.02209.x>.
- [3] Wixted, John T., y Shana K. Carpenter. «The Wickelgren Power Law and the Ebbinghaus Savings Function». *Psychological Science* 18, n.º 2 (febrero de 2007): 133-34. <https://doi.org/10.1111/j.1467-9280.2007.01862.x>.
- [4] Lindsey, Robert V., Jeffery D. Shroyer, Harold Pashler, y Michael C. Mozer. «Improving Students' Long-Term Knowledge Retention through Personalized Review». *Psychological Science* 25, n.º 3 (marzo de 2014): 639-47.
<https://doi.org/10.1177/0956797613504302>.