

## Natural computing applied to population genomics

Esta conferencia ha sido impartida por Oscar Lao, líder de *The Population Genomics group* en el CNAG-CRG. Nos ha explicado como funcionan los algoritmos naturales, que tienen una parte de *machine learning*, de una forma muy dinámica y entendible.

La hipótesis más aceptada de la evolución humana es que los humanos anatómicamente modernos salen de África y colonizan todo el mundo de una manera rápida. Los humanos anatómicamente modernos no estaban solos, también estaban los Neandertales y Denisovanos. A día de hoy, sabemos que hubo hibridación entre estos dos y los humanos anatómicamente modernos porque en nuestro genoma hay entre un 2 y un 4% de fragmentos provenientes del Neandertal. En Asia hubo una retroversión con los denisovanos y un 6% del genoma de los asiáticos proviene de los denisovanos.

Existe un algoritmo natural que intenta simular el funcionamiento de un cerebro humano porque tiene unas entradas de otras neuronas, a continuación hay variaciones de los inputs y finalmente envía la información a otras células. Esta estrategia se utiliza en el *deep learning*. Al algoritmo lo tenemos que entrenar y para ello se utilizan modelos, pero que tienen ruido porque de esta manera le va a costar aprender, pero cuando lo aprende lo hace bien. Si le ponemos un modelo muy sencillo y el patrón que tiene que buscar está muy escondido, el *deep learning* nos va a dar una respuesta errónea.

El *deep learning* que desarrolló el grupo de Oscar permitió corroborar dos hipótesis evolutivas que también fueron encontradas por otros grupos a partir de otras estrategias. La topología del *deep learning* que desarrolló el grupo de Oscar presentaba algunas limitaciones, por ejemplo, no era capaz de distinguir cuál de los dos siguientes modelos era mejor: (1) los Neandertales y Denisovanos hibridan a la vez con los europeos o (2) se formó un híbrido entre los Neandertales y Denisovanos y este fue el que hibridó con los europeos. Para solucionar este problema decidieron añadir información genómica del Neandertal y Denisovanos. Debido a que en los dos casos la información que queda en el genoma es distinta, en el primer caso los fragmentos adquiridos no estarán relacionados entre ellos, pero en el segundo caso sí que van a estarlo. Añadieron este nuevo punto de vista adaptando la metodología de reconocimiento en imágenes, asumiendo que a secuencia genómica es una imagen de una dimensión. El valor del estadístico que se calcula en las ventanas del genoma es distinto en función de donde provenga el genoma. Estos patrones se conocen como SFS. La adaptación no fue tan sencilla, también tuvieron que añadir un nuevo algoritmo natural para hacer un filtrado de SFS que le pasamos al *deep learning*, en concreto, este algoritmo era evolutivo.

Por lo tanto, el algoritmo final que desarrolló el grupo de Oscar tiene dos fases: (1) elegir que ventanas del genoma son interesantes con una estrategia evolutiva y (2) utilizamos la selección como output y la pasamos por el *deep learning* para obtener los output finales.

Hicieron una comprobación de los datos que se obtenían y vieron que, con las mejoras, el nuevo algoritmo sí que permitía distinguir las dos situaciones, no al 100%, pero muy próximo a ello, teniendo en cuenta que antes era un 50%, se ha mejorado mucho.

Con esta conferencia hemos aprendido que inferir parámetros evolutivos y demográficos es complejo y no solo hay una posible solución. En mi opinión, me ha parecido muy interesante porque había oído hablar del *machine learning*, pero no conocía ningún ejemplo donde se ha implementado.