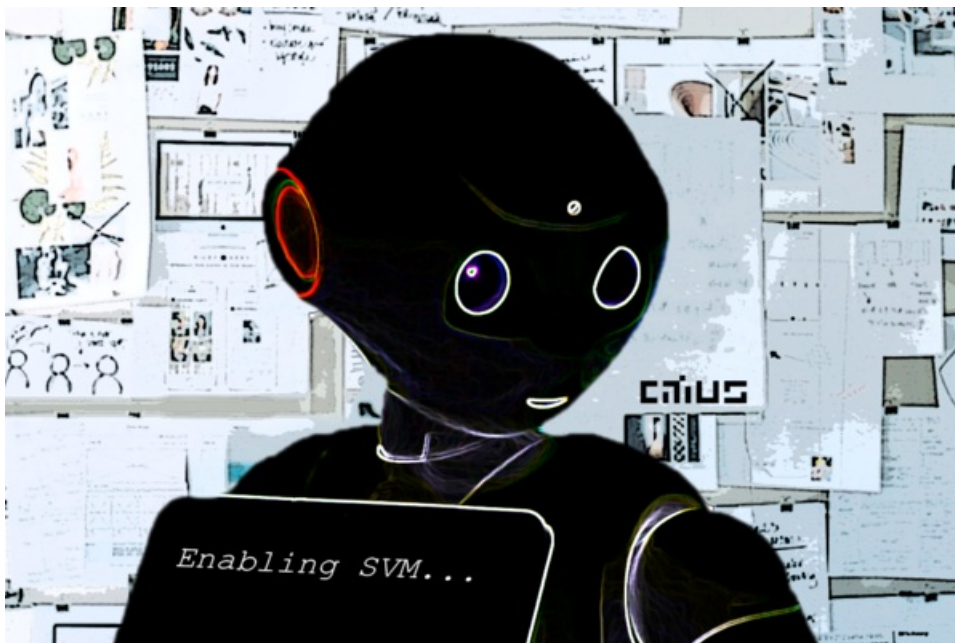


## Crean no CiTIUS un novo algoritmo de aprendizaxe automática ultrarrápido e de baixo consumo para o procesamento de grandes volumes de datos

xoves, 5 agosto, 2021 - 13:25

*A nova solución mellora en ordes de magnitude un dos métodos de referencia para a clasificación de datos mediante técnicas de Machine Learning. O traballo vén de publicarse na revista 'IEEE Transactions on Pattern Analysis and Machine Intelligence', a publicación líder entre as 140 revistas científicas máis citadas no ámbito de Ciencias da Computación-Intelixencia Artificial.*

A case ninguén se lle escapa xa que a Intelixencia Artificial (e máis concretamente a 'Aprendizaxe automática' ou *Machine Learning*), viven o seu momento de máximo esplendor. A investigación desenvolvida neste eido nos últimos anos, unida á enorme potencia de cálculo alcanzada polos computadores e a inxente cantidade de datos dispoñible para adestrar os algoritmos, revolucionou as nosas vidas e tódolos sectores da economía.



Nada diso tería sido posible sen a curiosidade e o esforzo da comunidade científica, que ao longo das últimas décadas contribuíu á progresiva sofisticación das técnicas de *aprendizaxe automática*, unha área de investigación cada vez máis precisa, que permite adestrar ás máquinas para que resolvan problemas moi diversos sen programalas especificamente para cada situación.

As estratexias propostas para logralo son moitas; unha das máis coñecidas e potentes hoxe en día son as denominadas 'Máquinas de Vectores de Soporte' (*Support Vector Machines, SVM*), creadas polos científicos Isabelle Guyon, Bernhard Schölkopf e Vladimir Vapnik na década dos 90. Trátase dunha achega de importancia capital, recoñecida entre outros galardóns co [Premio 'Fronteiras do Coñecemento' da Fundación BBVA en 2020](#).

Esencialmente, as Máquinas de Vectores de Soporte (SVM) son un método para clasificar conxuntos de datos, cunha precisión practicamente idéntica á humana; e nalgúns casos, mesmo maior. As SVM son un dos clasificadores con mellores prestacións na actualidade, que ten demostrado sobradamente a súa eficacia á hora de recoñecer datos de natureza diversa: desde textos, voces e rostros de persoas, ata células cancerosas ou usos fraudulentos dun cartón de crédito.

Con todo, non existe un método infalible nin mellor en tódolos casos, ou para todas as circunstancias. Así, as SVM reveláronse considerablemente lentas á hora de abordar problemas nos que o número de datos é moi grande, o cal resulta particularmente problemático á hora de traballar en contornas *Big Data*; por outra banda, o seu consumo de memoria resulta en ocasións inasumible, podendo invalidar, na práctica, este tipo de solucións.

Agora, unha investigación realizada no CiTIUS (*Centro Singular de Investigación en Tecnoloxías Intelixentes da Universidade de Santiago de Compostela – USC*), permitiu superar estas limitacións co desenvolvemento dun novo '*Clasificador Rápido de Vectores de Soporte*' (*Fast Support Vector Classifier, FSVC*), que presenta numerosas vantaxes respecto ao método estándar. En primeiro lugar, é moito máis rápido-entre 10 e 100 veces-que as aproximacións tradicionais. Ademais, este novo clasificador

opera con moita menos memoria, «grazas ao cal é capaz de desenvolver solucións óptimas con computadores moito menos potentes e custosos», explica **Manuel Fernández Delgado**, director da investigación.

Os investigadores do CiTIUS salientan esta cuestión como unha das achegas esenciais do traballo: «o aforro de memoria ten unha grande importancia», asegura **Ziad Akram**, investigador predoutoral do CiTIUS e primeiro autor do artigo, «xa que mellorando a eficiencia podemos resolver, con equipos moito máis modestos, problemas para os que normalmente precisaríamos un supercomputador». «Todo iso tradúcese nunha redución enorme do custe e o consumo de enerxía», subliña a súa colega **Eva Cernadas**, investigadora principal do centro e coautora do *paper*.

Outro dos artífices do traballo, **Senén Barro**, puntualiza que «unha das claves foi conseguir desenvolver unha solución analítica para o deseño de clasificadores, o que evita usar métodos iterativos de aprendizaxe sobre conxuntos de datos, a causa principal da ineficiencia computacional e de consumo de recursos da *Aprendizaxe automática*». O director científico do CiTIUS aclara que «con esta nova aproximación, é coma se puidésemos memorizar dunha vez un conxunto enorme de casos (de caras, por exemplo), sen necesidade de velas unha e outra vez ata que queden gravadas na nosa memoria». «A rapidez e o aforro en memoria e en capacidade de cómputo é enorme, o que supón un aforro en diñeiro e, aínda máis importante, en pegada de carbono», conclúe Barro.

#### Referencias

Z. A. Ali Hammouri, M. F. Delgado, E. Cernadas and S. Barro, "*Fast SVC for large-scale classification problems*". IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2021.3085969.

- Código en Octave: <https://persoal.citius.usc.es/manuel.fernandez.delgado/papers/fsvc>
- Cápsula en IEEE Ocean: <https://codeocean.com/capsule/2827247/tree/v1>