Joint Complexity with suffix and hyper-suffix trees

Philippe Jacquet[*1]

[1]Nokia - Bell Labs – - – France

Résumé

The joint complexity has been introduced for DNA comparison and has been successfully applied to natural text sampling and clustering. It consists into enumerating the common factors between two texts without taking into consideration languages, grammars and vocabularies. Therefore it is extremely fast to compute via the use of suffix trees. The larger is the joint complexity the closer are the texts in meanings. It provides results at least as accurate as machine learning but without any training and at unprecedented speed. It has been successfully experimented on Twitter flows. Furthermore, it can be precisely analysed via combinatorics and generating functions when the text generation process are Markovian with finite memory. The process of text clustering can even again been accelerated via the use of hyper-suffix trees which have been recently invented. We have called this new branch of Artificial Intelligence which uses combinatorics and fast algorithms, the "accelerated AI". Joint work with Gerard Burnside.

---

[*]Intervenant