
Interprétabilités des mesures d'accord inter-annotateur

Jean-Yves Antoine*¹ and Matéo Malidin*[†]

¹Laboratoire d'Informatique de l'Université de Tours (LIFAT) – Polytech'Tours, Université François Rabelais - Tours : EA6300 – 64, Avenue Jean Portalis, 37200 Tours, France

Résumé

Le machine learning et le deep learning font souvent appel à des données qui ont été obtenues par annotation manuelle. Le processus d'annotation étant le plus souvent subjectif, la qualité des données est évaluée par des mesures de fiabilité reposant sur une mesure d'accord inter-annotateurs. Une des limites principales de ces mesures d'accord est leur manque d'interprétabilité. Des recherches antérieures ont montré la pertinence de relier la mesure d'accord inter-annotateurs avec le taux de stabilité d'une annotation de référence afin d'en établir une interprétation objective et générique. Cette présentation résumera les expériences qui ont été menées pour étudier plus finement la corrélation entre une mesure d'accord (le Kappa de Cohen) et le taux de reproductibilité de l'annotation de référence, en travaillant sur des données réelles et synthétiques générées spécifiquement afin de faire varier artificiellement l'accord entre les annotateurs. Nos résultats montrent qu'il existe une forte corrélation (non linéaire) entre accord et taux de reproductibilité. Toutefois, cette corrélation varie légèrement suivant les corpus considérés, ce qui nous interroge sur d'autres facteurs impactant l'interprétation de ces mesures de qualité des données.

Mots-Clés: Qualité des données, Annotation, Accord inter, annotateur, Interprétabilité, TAL

*Intervenant

[†]Auteur correspondant: malidin.mateo@gmail.com