

Sujet PLDAC

Review spam & classification d'auteurs

Vincent Guigue & Nicolas Baskiotis
vincent.guigue@lip6.fr, nicolas.baskiotis@lip6.fr

Sujet ouvert à un binôme

Les revues sur le web participatif constituent une source particulièrement particulièrement riche : ces données sont disponibles quasiment à l'infini, elles sont multilingues, classée par thématiques, étiquetées en sentiments (et parfois même sur différents aspects)... Mais ce système fait l'objet d'attaque de plus en plus nombreuses : le *review spam* (génération de fausses revues). Ces revues sont très difficiles à détecter car la plupart du temps, elles sont écrites par de véritables personnes...

Dans un projet précédent, nous avons étudié une stratégie de type PageRank [Wang 2011] attribuant une note de confiance aux revues, aux produits et aux utilisateurs : un spammeur est suspecté d'avoir envoyé plusieurs faux messages, un produit est rarement attaqué de manière isolé... Toutes ces hypothèses se renforcent les unes les autres pour aboutir à un système fiable. Cette stratégie est intéressante mais présente un défaut majeur : elle ne tient pas compte du contenu des revues. L'idée est de construire un nouveau système tirant parti de la proximité de style entre différentes revues pour détecter les auteurs publiants sous de multiples pseudos. Au cours des années passées, nous avons réalisé un projet de classification d'auteurs sur lequel les étudiants pourront s'appuyer.

Le but est donc de construire un système basé à la fois sur la dynamique et le contenu et d'exploiter les nouvelles bases de données annotées en spam pour construire un nouveau système.

Les développements se feront en python. Ils utiliseront d'abord `sklearn` puis éventuellement des outils de deep learning comme `pytorch` si les premières étapes sont dépassées rapidement.

[Wang 2011] *Guan Wang, Sihong Xie, Bing Liu, and Philip S. Yu, Review Graph based Online Store Review Spammer Detection, ICDM'11*