4-IF-FD – Projet de fouille de données

Fouille de données du Web : Etude et prédiction de l'audience d'une plateforme de streaming

Mehdi Kaytoue – Jean François Boulicaut – 2013/2014

Contexte

George Bernard Shaw once wrote that "We don't stop playing because we grow old, we grow old because we stop playing..." Enjoying video games at a professional level is not a young boy dream anymore and the best evidence is the amazing evolution of electronic sports over the last decade. Similarly to traditional sports, electronic sports attract a vast community of professional players (pro-gamers), teams, commentators, sponsors, and most importantly, spectators and fans. Indeed, a recent social study has shown that video game players prefer watching pro-gamers playing, rather than playing themselves.

The main difference with respect to traditional sports lies in the fact that the vast majority of the events are only online and an important remark is that members of the community are acquainted with social networks such as Facebook or Twitter and web platforms YouTube. As a consequence, a new type of social community is emerging, very active on several web social platforms and of a particular interest for the social network research community.

***Online live video streaming of video games**, or social TV, now attract millions of spectators on a daily basis. This success is mainly visible on Twitch.tv, a live video streaming platform. Typically, major tournaments are broadcast, but generally a single player broadcast his games, chats, explains his game style and gives advices, which finally induces new kinds of relationships between him and his spectators.*

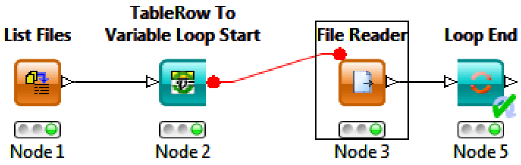
Concepts principaux (mais non limité à !)

- **Exploration et description de données**
- **Clustering**
- **Motifs fréquents et règles d'associations**
- **Régression linéaire et prédiction**
- **Visualisation**
- **Knime, Sci-Kit Learn (python)**
- **XML**

Objectifs et résultats attendus

Your goal is to give a longitudinal characterization of this new community by analyzing Twitch audiences. Previously, you crawled the list of active live video streams along with their respective number of viewers every five minutes from September 29th, 2011 to January 09th, 2012. Your data analysis should enable (i) to characterize video streams qualitatively (identifying the games and the player location) and quantitatively through their viewer counts, durations, and audience, (ii) to early predict the audience of a stream, and finally (iii) to rank the most popular players. These results are of major interest for all actors of this community. For example, popularity is key in a pro-gamer career, strongly influencing his revenues (sponsors, invitations to tournament with prizes and advertisement revenues while streaming).

Conseils : N'oubliez pas d'installer des extensions de KNIME, et de le mettre à jour. Par exemple, pour décompresser un zip, ou encore utiliser XPATH sur les fichiers XML. Pour traiter des dossiers de fichiers de données, utiliser les méta-nœuds, comme ci-contre par exemple.



Information : le jeu complet se trouve sur http://intoweb.loria.fr/MSND_WWW_DATA/rough/