



INF8808 - VISUALISATION DE DONNÉES

PROJET DE SESSION

Plan

Groupe : 9

[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

22 Février 2019

1 Option : Starcraft II

1.1 Jeu de données

La base de données de Aligulac contient de l'information sur une très grande quantité de parties professionnelles jouées pendant, comme les joueurs impliqués, l'évènement, la cote des joueurs, les groupes des tournois, les équipes.

D'autres données qui sont disponibles concernent de l'information plus précise sur des parties jouées, selon plusieurs niveaux différents de joueurs. On y retrouve de l'information sur l'âge des joueurs, le nombre d'heures jouées par semaine, le total des heures jouées, le nombre d'actions par minute durant la partie, et d'autres différentes méta données sur la partie qui peuvent être d'intérêt. Cet ensemble de données était à l'origine constitué dans le but de faire de l'apprentissage machine.

Il est également possible d'obtenir des informations générales sur le jeu à partir du site officiel de Blizzard, ainsi que sur d'autres sites externes populaire tels que Liquipedia, ect.

Finalement, il existe un package python permettant de lire un fichier "replay" de Stacraft 2 et d'en sortir de nombreuses méta données. Nous avons testé ce package, et il fonctionne correctement. Nous pourrions alors choisir des parties spécifiques et en déduire les métas données.

Liens

1. Aligulac : <http://aligulac.com/about/db/>
2. Données sur les parties : <http://archive.ics.uci.edu/ml/machine-learning-databases/00272/>
3. Package python : <https://media.readthedocs.org/pdf/sc2reader/latest/sc2reader.pdf>

1.2 Questions possibles

1.2.1 Public cible : Personnes qui sont intéressées à Starcraft II, mais qui ne connaissent pas beaucoup le jeu

Contexte général

1. Quelle est la popularité de StarCraft II (nombre de joueurs par rapport à d'autres jeu et autres jeu professionnels) ?
2. Quelle est la valeur de la communauté professionnel ? (Quantité d'argent que ça représente (ou salaire/revenu) comparé à d'autres jeu et comparé à une référence plus intuitive (e.g. un autre sport professionnel))
3. Quelle importance a apporté le jeu dans la scène des sports électroniques ?

Jeu en tant que tel

1. Comment se comparent les unités de certaines races selon des statistiques différentes (attaque, défense, porté, dégâts par seconde, dégâts de masse (porté des dégâts de masse), vitesse de déplacement)
2. Qui gagne dans une escarmouche (telle unité vs. telle autre) ?

3. Quelles types de compositions d'unités sont-elles les plus populaires pour chaque race ?
4. Que veut dire le vocabulaire compliqué employé par les joueurs et les commentateurs ?
5. Quelles sont les dépendances entre les bâtiments, les unités, les upgrades, etc ?

1.2.2 Public cible : Joueurs expérimentés

1. Quelles upgrades (défense, attaque, bouclier, vitesse, range...) sont des choix décisifs / sont de meilleurs choix ? Dans quelle situations peuvent-elles devenir importantes / devenir un meilleur choix qu'une autre ?
2. Quels sont les build orders les plus agressifs (temps avant une attaque) ? Les plus intensif au niveau de l'APM (qui demandent un haut niveau de macro/micro) ? Les plus prudents (qui permettent de défendre un grand nombre d'attaques en début de partie) ?
3. Comment se compare l'APM et l'EMP (Effective Actions per Minute) pour des joueurs débutants, moyen, avancés, experts ? Selon les races ?
4. Dans les parties professionnelles, quand est-ce que le APM est le plus élevé ? Pourquoi ? (e.g. juste spam ?)
5. Comment se comparent les différentes stratégies entre les races ? (expansion timing, nombre de bases, nombre d'unités, nombre de bâtiments, required APM etc.) AMP required for executing different builds / strategies effectively ? (some builds or unit compositions are more APM intensive)
6. Informations sur les joueurs professionnels (pays (korea vs le reste du monde), âge, âge lorsqu'il a commencé à jouer, race, pourcentage de victoires, argent gagné).

2 Option : Traces de kernel Linux

2.1 Jeu de données

N'importe quelle trace de kernel Linux peu servir comme jeu de données. Il serait nécessaire par contre de générer des traces qui représentent des situations problématiques intéressantes afin de pouvoir visualiser un cas intéressant.

2.2 Questions possibles

1. Quelle est la source d'un problème que j'ai trouvé ? Pourquoi est-ce la source du problème (la cause) ?
2. Y a-il des problèmes importants de latence ? Où ? Quel est le contexte de cette latence, les événements qui se sont passés autour de l'événement (afin de mieux comprendre la source de la latence)
3. Est-ce que le travail est bien distribué entre les cores du système ?
4. Problèmes d'obtention des ressources ? Y a-t-il des inversions de priorités ? (Systèmes temps-réel)

5. Pourquoi le système a-t-il manqué le deadline ? Problèmes d'ordonnancement ? (Systèmes temps-réel)