

# IA TP2 : Pattern Mining

## Rapport

Mathias Dugue

Valentin Hulot

# Extraction et Transformation des données

L'extraction et transformation des données à été réalisée avec panda.  
Le processus et les résultats sont disponibles sur le dépôt de l'Istic.

**Disponibles sur le [gitlab](#):**

- Le jupyter notebook en .ipynb (TP2\_HS\_data.ipynb)
- Le jupyter notebook en .pdf (TP2\_HS\_data.pdf)
- Les datasets pour la partie A et B

**Les datasets fournis sont les suivants :**

- outputWithPowers.txt  
(transaction database pour LCM <sup>1</sup>avec cartes et AVEC pouvoirs des héros)
- outputWithoutPowers.txt  
(transaction database pour LCM avec cartes et SANS pouvoirs des héros)
- outputSequentialWithPowers.txt  
(sequence database pour CloSpan <sup>2</sup>avec cartes et AVEC pouvoirs des héros)
- outputSequentialWithoutPowers.txt  
(sequence database pour CloSpan avec cartes et SANS pouvoirs des héros)

---

<sup>1</sup> Linear Closed Itemset Miner

<sup>2</sup> Closed Sequential Pattern Mining

# Analyse des résultats

## Partie A : Recherche de decks types

### Questions

#### **Que représente un itemset fréquent fermé ici ? Pourquoi est-ce potentiellement intéressant ?**

Les CFI (Closed Frequent Itemsets) représentent un ensemble d'items qui apparaissent souvent ensemble dans une série d'un jeu de données.

Ici, ils représentent les cartes fréquemment jouées ensembles au sein d'un même deck.

C'est potentiellement intéressant pour trouver des relations entre ces objets, comme :

- Suggérer la prochaine carte (ou plus) après la sélection de la première.
- Améliorer l'arrangement de ces items dans un catalogue, un rayon
- Prédire un comportement utilisateur  
(si carte A et B sont choisies, alors le joueur prendra sûrement C)

#### **Y a-t-il des itemsets fréquents fermés « triviaux » qui ne nous apprennent pas grand-chose ?**

Les CFI triviaux sont les items seuls, cela nous donne leur fréquence d'apparition, mais ne nous donne pas d'autres indications.

Ils sont utilisés pour générer les CFI de plus grande taille.

Il y a aussi les CFI contenant les pouvoirs de classe parce que ce sont des items très utilisés en partie que ce soit pour vider son mana ou pour un combo particulier et le fait que ce soit une mécanique de jeu "universelle" utilisable à chaque tour peu chère ce qui garantit leur présence dans une partie.

#### **Avez-vous trouvé des itemsets fréquents particulièrement intéressants ?**

##### **1. CFI de taille 1**

En ce qui concerne ces objets, on peut formuler quelques hypothèses. Les cartes qui sont très jouées (aka. Score #SUP élevé) pourraient l'être car elles sont très fortes, polyvalentes, ou encore disponibles dès le départ dans le jeu.

##### **2. Cartes fréquentes par classe**

On peut observer de nombreux itemsets contenant les sorts des Héros.

Par exemple, certains CFI contiennent LesserHeal. On peut donc en déduire que ces cartes sont souvent jouées avec la classe Prêtre.

Ex :

Class : Priest (lesserHeal)

Number of decks : 56 (Obtained with LesserHeal #Sup: 56)

Title	Sup #	%
LesserHeal PowerWord:Shield	39.0	70
LesserHeal NorthshireCleric	38.0	68
...		
LesserHeal NorthshireCleric PowerWord:Shield	28.0	50
LesserHeal TwilightGuardian WymrestAgent	27.0	48
...		
LesserHeal PowerWord:Shield TwilightGuardian WymrestAgent	22.0	39
LesserHeal PowerWord:Shield TwilightGuardian TwilightWhelp WymrestAgent	19.0	34

Si on analyse les decks pour la classe prêtre seulement, on peut voir différentes occurrences pour les mêmes cartes.

Les deux premières lignes (2 items) montrent que ces cartes sont jouées dans la majorité des cas avec un deck prêtre.

On peut ensuite voir sur les lignes avec 3 items que certaines cartes apparaissent souvent ensemble.Ex:

- NorthshireCleric & PowerWord:Shield
- TwilightGuardian & WymrestAgent

Ces cartes sont potentiellement des combos, des cartes qui fonctionnent bien ensemble. Il faudra étendre l'analyse pour voir si elles sont jouées à la suite.

On peut parfois trouver des ensembles plus conséquents dans le deck. Ils apparaissent à des fréquences moins élevées. Il faudrait voir à partir de quel seuil filtrer, en fonction de l'objectif à remplir/monttrer grâce à ces résultats.

## (Bonus) Partie B : Séquences et règles séquentielles

### **Que représente un itemset fréquent fermé ici ? Pourquoi est-ce potentiellement intéressant ?**

Les CFI (Closed Frequent Itemsets) représentent un ensemble d'items qui apparaissent souvent ensemble dans une série d'un jeu de données.

Ici, ils représentent les cartes fréquemment jouées le tour d'après la carte la précédant dans l'itemset.

Ces données sont potentiellement intéressantes car elles permettent de trouver des cartes souvent jouées en réponse à une autre carte et donc de trouver de possibles contres.

De plus, si l'itemset est plus long que 2 cartes ceci permettrait de découvrir des séquences de jeu standard avec des sortes de "contre du contre" ou alors des réponses au contre adverse qui pourraient être intéressants de rajouter.

Il faut tout de même faire attention à la présence de cartes "standard" ou à bas coût qui sont très souvent jouées pour terminer la barre de mana plus que pour contrer une certaine carte ou d'autres genre de situations avec des cartes standard.

### **Y a-t-il des itemsets fréquents fermés « triviaux » qui ne nous apprennent pas grand-chose ?**

Il y a tout d'abord les itemsets de taille 1, qui représentent (encore) les cartes jouées souvent, mais ici comme les données sont organisées tour par tour et non partie par partie avec usage unique, cela permet d'avoir plus d'informations sur la fréquence d'utilisation de la carte. Ceci permettrait de déduire que c'est potentiellement une carte à faible coût, qui serait utilisée souvent en début de partie ou alors pour terminer la barre de mana.

Après dans le cadre du dataset gardant les pouvoirs des héros, l'utilisation omniprésente de ces derniers reflètent encore leur caractère universel, à coût faible et utilisable une fois par tour.

Finalement, les itemsets de type Carte A | Carte A, reflètent le caractère non attaché à une classe d'une carte comme elle sont souvent présentes dans des decks très divers et sont hypothétiquement avec un coût faible car souvent les cartes avec un coût un peu plus fort sont spécifiques aux classes.

### **Avez-vous trouvé des itemsets fréquents particulièrement intéressants ?**

Les itemsets de taille 2 où les cartes à bas coût et les pouvoirs de héros ne sont pas présents sont ceux les plus intéressants pour cette taille, car on peut en déduire que le tour joué en contre à la première carte tourne autour de cette carte, qui a donc potentiellement été jouée stratégiquement pour contrer la carte précédente.

Pour les itemsets de taille supérieure à 2, ils permettraient de voir des stratégies sur le long terme ou des séquences de contres permettant de comprendre comment adapter son deck en fonction de la classe adverse afin de mieux jouer le match-up.

Cependant ils sont tout de même assez rares et des cartes à bas coût peuvent venir affaiblir leur intérêt.

Avec plus de connaissances, nous pourrions “bannir” les cartes à bas coût de notre recherche afin d’obtenir des séquences de jeu beaucoup plus intéressantes.

## Points d’amélioration

L’utilisation d’un jeu de données contenant l’issue de la partie (win ou loose) permettrait de considérer seulement les decks/cartes les plus efficaces.

Des visualisations supplémentaires auraient pu être réalisées sur la data de sortie de LCM et CloSpan.