

Evaluation des dossiers sur les mathématiques du Big Data

Les « 12 travaux d'AstéRix »

Daif Hakim

31/01/2021

Marche Aléatoires

1. Critères d'évaluation

Dans ce dossier, nous allons parcourir et faire une évaluation des travaux de mes camarades en mathématiques du Big Data:

Nos critères seront :

- 1. Visuel et l'organisation du dossier:** Es ce que la mise en page est propre, agréable à lire ou non ? Le travail est-il organisé (chapitres, sections, sous-sections, Parties)
- 2. Qualité du latex et des formules mathématiques :** Es ce que les formules sont claires ou pas ? Es ce que le dossier contient le code Latex ou le rapport sous format Rmd ?
- 3. Compréhension de l'idée générale:** Les Auteurs arrivent-ils à transmettre l'idée générale ?
- 4. Explication et compréhension des formules mathématique et des concepts fondamentaux:** Es ce que les formules et les notions sont bien expliquées ?
- 5. Difficulté et originalité du thème :** Le sujet abordé est-il original/difficile ou pas ?

2. Lien vers le document commenté

l'un des travaux que nous allons analyser dans ce dossier est celui de WILLIAM et MARKO sur le thème qui porte sur les Marches Aléatoires.

Les liens vers leurs Github:

<https://github.com/WilliamRbc/PSBX/tree/main/MARCHE%20ALEATOIRE>

Le dossier en question est **MARCHE ALEATOIRE.Rmd**

Ou bien en format PDF : **MARCHE-ALEATOIRE-word.pdf**

3. Synthèse de la présentation

La marche aléatoire est un modèle mathématique qui permet de prédire le mouvement d'un système en fonction de son point de départ. les auteurs donnent tout d'abord une définition très simple de la marche aléatoire, ils expliquent également son concept général avec de simples exemples Les auteurs ont tenté de savoir si lors de l'étude des marches aléatoires, le système revient à son point de départ, c'est-à-dire s'il revient à son origine, cependant on comprend que la marche aléatoire peut être récurrente ou transiente.

WILLIAM et MARKO, afin d'expliquer la marche aléatoire isotrope ils ajoutent l'exemple d'une personne qui est sur un escalier et qui tire pile ou face pour décider si le prochain pas sera vers le haut ou vers le bas, le but étant de construire le test de Bernoulli avec issues équiprobables.

les auteurs ont produit 3 simulations de cette expérience et ils ont tracé à chaque simulation le graphe qui correspond.

les auteurs reviennent sur exemple déjà cité dans l'introduction qui porte sur le fait que Google utilise la marche Aléatoire afin d'expliquer le Modèle de marche aléatoire et l'indice PageRank.

Dans cette partie les auteurs vont tenter d'expliquer le principe de PageRank, pour cela ils font un rappel bases mathématiques sur les chaînes de Markov et le théorème d'une chaîne dite ergodique.

4. Explication des formules

Ce qui faut retenir des formules utilisées par les auteurs pour expliquer la marche aléatoire :

On appelle marche aléatoire associée à la suite de variables aléatoires $(S_n)_{n \geq 1}$ où S_n est défini par

$$S_n = X_1 + \dots + X_n.$$

sachant que (X_i) une suite de variable aléatoires indépendantes à valeurs dans $\{(\pm e_1, \dots, \pm e_d)\}$

Cette première formule nous définit d'une manière très simplifiée la marche aléatoire.

les auteurs nous rappellent les chaînes de Markov puis indiquent la différence entre une distribution stationnaire et ergodique et tout ça dans le cadre du Modèle de marche aléatoire et PageRank

la distribution Px est stationnaire si et seulement si Px est un vecteur propre de P^t de la valeur propre 1:

$$Px = P^t * Px$$

la distribution est ergodique si elle converge vers une même chaîne

$$\forall P(X_0), \lim_{t \rightarrow +\infty} P(X_t) = P(X_\infty)$$

Pour mettre en place le PageRank, google utilise Les chaînes de Markov, cependant, cela ne suffit pas pour pouvoir définir la notoriété de la page à cause du problème de culs-de-sac autrement dit la chaîne n'est pas irréductible

Solution : introduction d'une probabilité α de reset de la marche aléatoire

$$P_{\text{pagerank}} = (1 - \alpha) \cdot P + \alpha \cdot \frac{1}{n} \cdot J \text{ avec } \forall i, \forall j, J_{ij} = 1 \text{ Typiquement } \alpha = 10\%$$

5. Evaluation

1. Visuel et l'organisation du dossier: Le visuel est pas mal , mais le travail n'est pas vraiment organisé vu qu'on a du mal à comprendre l'enchaînement des idées, (on sent que les idées ne sont pas reliées, dès fois on a l'impression que les titres des parties ne correspondent pas aux contenus et cela est sûrement du au travail non complet sur quelques parties notamment la partie marche aléatoire Isotrope et celle du PageRank

2. Qualité du latex et des formules mathématiques : Déjà on trouve le rapport fait en latex ce que j'apprécie bien cependant on constate qu'il y a des erreurs dans le code latex vu qu'il a quelques formules qui ne sont pas claires notamment celles de la partie PageRank et celles de la partie marche aléatoire isotrope

An constate aussi l'utilisation des captures d'écrans de quelques formules cela montre que les auteurs ont quelques difficultés dans l'utilisation de latex (en ce qui concerne les formules)

3. Compréhension de l'idée générale : On l'impression que les auteurs maitrisent les notions basiques, mais ils ont du mal à entrer dans les détails, cela on l'a constaté à travers la partie du PageRank

4. Explication et compréhension des formules mathématique et des concepts fondamentaux : Au début on trouve des formules claires et bien concises notamment pour la définition de la marche aléatoire, mais à la fin, et vu qu'ils ont raté leur code latex dans quelques parties on a eu du mal à les comprendre quelques erreurs à corriger aussi dans le code latex comme par exemple: $\langle! -\forall P(X0), \lim_{t \rightarrow +\infty} P(Xt) = P(X_\infty) -\rangle$ au lieu $\langle! -P(X_\infty)$ lorsque $t\beta + \infty -\rangle$

5. Difficulté et originalité du thème : le sujet abordé est original : En termes de difficultés, les notions introduites (exposées) sont basiques, mais si les auteurs ont développé encore plus les différentes parties et notions le sujet sera encore plus difficile

6. Conclusion

En conclusion, on trouve que le sujet est très intéressant et original j'apprécie trop le choix de WILLIAM et MARKO , cependant l'organisation de leur travail n'est pas à la hauteur de leur choix, les auteurs ont du mal à entrer dans les détails ce qui complique la compréhension générale de ce dossier. Il est nécessaire aussi aux auteurs de travailler un peu plus leurs codes latex pour que les formules soient un peu plus claires. Le concept des marches aléatoires nécessite un niveau un peu avancé de la compréhension des notions mathématiques , notamment en ce qui concerne les PageRank, malheureusement ils n'ont pas assez détaillé leur travail, manque des exemple pratiques pour une meilleure compréhension , manque de notions mathématiques avec des explications détaillées .