

Compte rendu : Calcule d'un score robuste **ET** équitable



Realisé par :

- Ismail EL MOUFLIH
- B M
- Alpha Oumar KANTE
- Aberrahim BOUKIOUD

Introduction:

L'objectif de ce TP est de proposer un mécanisme de ranking qui permet de calculer un score robuste et équitable en testant les différentes attaques possible, afin de vérifier si la solution proposé est bien robuste.

1) Dans un premier temps on a proposé un calcul de rang à partir de la moyenne des scores donnés, en se basant sur la formule suivante:

$$R = (r_1 + r_2 + r_3 + r_4 + \dots + r_N) / N$$

2) la deuxième solution proposée est le rang pondéré, qui fonctionne avec la formule suivante

$$R = (r_1 * t_1 + r_2 * t_2 + r_3 * t_3 + r_4 * t_4 + r_5 * t_5) / (t_1 + t_2 + t_3 + t_4 + t_5)$$

r : le score de chaque étoile

t : le nombre total de votes

3) la troisième proposition, si un client note plusieurs fois le même chauffeur, soit **i** la variable associée à chaque client et **n** total de clients avec **j** la variable associée au numéro de rang d'un client particulier.

$$R = \{ \text{somme}(R_{1j}) / j + \text{somme}(R_{2j}) / j + \text{somme}(R_{3j}) / j + \dots + \text{somme}(R_{nj}) / j \} / i$$

Les trois premières propositions sont plus efficaces pour les systèmes de grands volumes, ainsi que l'aspect de robustesse se pose toujours dans ces trois méthodes pour cette raison qu'on a proposé une quatrième méthode plus efficace et robuste utilisée par plusieurs systèmes de ranking.

4) la quatrième proposition est la méthode **Bayesian**

Après des heures de recherche nous avons découvert plusieurs autres méthodes, une des méthodes qui répond plus à nos besoins est la méthode de Bayesian.

Cette méthode nous permet de définir un système de ranking avec un mécanisme de confiance, qui permet de créer un score en fonction de la valeur de confiance affectée pour chaque chauffeur

Bayesian :

In 1770s, Thomas Bayes introduced 'Bayes Theorem'. Even after centuries later, the importance of 'Bayesian Statistics' hasn't faded away. In fact, today this topic is being taught in great depths in some of the world's leading universities.

Exemple :

How can we deal this issue ?

DVD Player	5 stars	4 stars	3 stars	2 stars	1 star	Total
Panasonic	5	3	3	0	0	11
Sony	18	9	5	2	3	37
Phillips	23	15	11	5	13	67
Curtis	18	14	2	6	11	51
Toshiba	19	10	10	4	11	54
Total	83	51	31	17	38	220

- Knowing total number of reviews gives us prior knowledge
- Find overall average
- Use overall average as Backup

so the idea is that the more ratings ther are for given individul product, the more trustworthy in the individual average should be relative to the overall.

- Bayesian ranking formula

$$\frac{\text{Overall Num.} \times \text{Overall Avg.} + \text{Individual Num.} \times \text{Individual Avg.}}{\text{Overall Num.} + \text{Individual Num.}}$$

Simple example :

- overall : reviews : 100, avg:2
- iniviua : reviews : 5, avg : 4

DVD Payers : 220 ratings ; 3,564 overall avg

DVD Player	No. Ratings	Naïve		Bayesian adj.	
		Ranking	Raw avg. Rating	Ranking	Rating
Panasonic	11	1	4.182	2	3.593
Sony	37	2	4	1	3.626
Phillips	67	3	3.448	4	3.537
Curtis	51	4	3.431	3	3.539
Toshiba	54	5	3.407	5	3.533

la valeur Bayesian calculée==>

Panasonic : 3.593

Sony : 3.626

Phillips : 3.537

Curtis : 3.539

Toshiba : 3.533

Ordering : sony , Panasonic , Curtis Phillips Toshiba

Conclusion

Bayesian average ratings are an excellent way to sort items with up-votes and down-votes, and lets us incorporate a desired level of caution directly into the model. It is eminently “hackable” in the sense of affording opportunities for adjusting the model to accommodate new features (prior beliefs, belief decay) without violating the model’s consistency. As long as we make a judicious choice of belief structure (namely a beta distribution), it is feasible to compute.

As with other hackable systems, it is possible to take the Bayesian framework and produce something utterly useless. But as long as you set aside time for a little experimentation, I think you’ll eventually arrive at a sorting method that helps people find the good stuff quickly.