

Comment prédire quel résultat a le plus de chances d'être cliqué dans les SERP ?



Par Guillaume et
Sylvain Peyronnet

| | | |
|------------------|------------------|----------------------|
| Domaine : | Recherche | Référencement |
| Niveau : | Pour tous | Avancé |

Bien sûr, le but et l'objectif d'un moteur de recherche est de proposer les résultats les plus pertinents répondant à une requête donnée. Pourtant, si ces liens ne sont pas cliqués, le moteur a échoué dans sa (re)quête. Il est donc important de tester comment une SERP (page de résultats) est analysée par l'internaute et de prendre en compte deux notions importantes : la pertinence et l'attractivité du résultat renvoyé. Cet article décrit des travaux de recherche dans ce domaine, travaux qui ont certainement un impact sur la façon dont nous proposons nos pages à des outils comme Google.

L'article de ce mois présente un résultat scientifique qui a été présenté à la conférence WSDM 2016 (*Web Search and Data Mining*) qui a eu lieu à San Francisco en février 2016. La précision est importante, car WSDM est, avec la très prestigieuse conférence WWW, l'une des deux rencontres les plus importantes des acteurs du search. Cette année, Jeff Dean (le spécialiste du deep learning chez Google), Yoelle Maarek (VP recherche de Yahoo !) ou encore Lars Backstrom (responsable du fil d'actualités chez Facebook) faisaient partie des orateurs invités.

C'est donc l'endroit pour découvrir les techniques que les moteurs mettront probablement en place dans les prochains mois, et dans cet article, nous allons vous présenter un travail qui a pour objectif de re-ranker, c'est-à-dire de reclasser les SERP en utilisant les caractéristiques des snippets pour prédire ce qui a toutes les chances d'être cliqué par les utilisateurs d'un moteur de recherche. Ce travail est dû à Jiepu Jiang et James Allan, et est décrit dans l'article [1].

Prédire la probabilité d'être cliqué, pour quoi faire ?

On pourrait oublier facilement l'objectif d'un moteur de recherche et penser qu'il ne sert qu'à classer quelques pages web (10 par exemple) dans l'ordre de décroissance de pertinence supposée par rapport à une requête. Hors, il n'en est rien : le but d'un moteur de recherche est de faire en sorte qu'un résultat pertinent soit vu par l'utilisateur. De ce point de vue, si le moteur met en avant une page très pertinente, mais que le lien vers cette page n'est pas cliqué, alors l'objectif n'est pas rempli, c'est un échec pour le moteur. *A contrario*, mettre en avant une page non pertinente via un lien qui sera cliqué est tout autant un échec pour le moteur.

Pour résoudre ce problème de correspondance entre les liens présentés, la pertinence des pages associées et le clic sur ces liens pour accéder à ces pages, plusieurs angles d'attaque peuvent être considérés. On peut par exemple travailler sur l'amélioration de la qualité des snippets, permettant aux utilisateurs de prendre une décision plus informée avant de cliquer ou non. Cette approche est cependant très peu efficace : Yilmaz et ses coauteurs (voir l'article [2]) ont en effet montré que la probabilité de cliquer sur un mauvais résultat restait de 49% dans ce contexte. On peut également éviter le problème en répondant directement à la requête (si il n'y a plus de nécessité de cliquer sur un résultat, on élimine le problème des clics malheureux). Bien sûr, cette astuce ne fonctionne que pour des requêtes directement informationnelles comme par exemple « quelle est la date de l'épreuve de philosophie du bac 2016 ? ».

L'idée des travaux de Jiang et Allan est finalement assez simple : quand deux pages sont de pertinences raisonnablement similaires, on déclassera la page qui a un plus haut taux d'erreur de clics ou de skips (nous verrons un peu plus loin ce que sont ces taux). L'idée est simple, mais la réalisation l'est moins : pour implémenter cette idée, il faut en effet analyser de nombreuses caractéristiques des snippets présentés par un moteur de recherche, et faire une analyse de corrélation entre ces caractéristiques et la satisfaction des utilisateurs.

Les notions clés des travaux de Jiang et Allan

Pour quantifier la qualité d'une page de résultats d'un moteur de recherche, Jiang et Allan utilisent deux notions complémentaires : la pertinence, qui est la notion classique pour exprimer qu'une page satisfait le besoin informationnel sous-jacent à une requête, et l'attractivité, qui est la suspicion de pertinence que peut avoir un utilisateur en voyant le snippet d'un résultat donné. Bien évidemment, pour évaluer la pertinence, il faut avoir cliqué sur le résultat correspondant, et donc une page pertinente a tout intérêt à être également attractive. Au final, la pertinence et l'attractivité sont souvent corrélées, mais peuvent être en désaccord (page pertinente et non attractive, ou page attractive mais non pertinente).

En regard de ces deux notions, les auteurs de l'article identifient deux types d'erreurs totalement complémentaires : l'erreur de clic (« click error » en anglais) et l'erreur de saut (« skip error » en anglais). Une erreur de clic a lieu lorsqu'un utilisateur clique sur un résultat qui va s'avérer non pertinent tandis qu'une erreur de saut arrive lorsqu'un utilisateur ne va pas cliquer sur un résultat qui aurait été pertinent pour lui. Dans la suite, on va noter p_1 la probabilité de cliquer sur la page P1, et r_1 la pertinence de cette page.

En étudiant les propriétés mathématiques associées à ces notions, Jiang et Allan proposent plusieurs règles de classement pour obtenir des SERPs de meilleure qualité :

1. Si $p_1 > p_2$ et $r_1 > r_2$ alors on classe P1 avant P2.

C'est le cas le plus facile : si P1 est plus pertinente et plus attractive, alors on va classer P1 en premier, c'est totalement naturel.

2. Si $r_1 > 0$ et $r_2 \simeq 0$ alors on classe P1 avant P2.
Une page pertinente est toujours proposée avant une page non pertinente.
3. Si $r_1 \simeq r_2 \simeq 0$ et $p_1 < p_2$ alors on propose P1 avant P2.
En effet, P1 a moins de chance d'être cliquée, donc on la propose en priorité vu qu'aucune page n'est pertinente.
4. Dans les autres cas, on propose en premier la page qui a la plus grande valeur de $r * p$. Ainsi, si $r_1 * p_1 < r_2 * p_2$ on proposera d'abord p_2.

A la lecture de ces règles, une question doit vous venir aussitôt à l'esprit : comment fait-on pour les vérifier, sachant qu'on ne peut connaître la valeur de la probabilité de clic qu'une fois que les utilisateurs auront cliqué ? C'est le problème de l'œuf et de la poule : pour calculer la probabilité il faut avoir fait un classement visible des utilisateurs, et pour faire le classement on a besoin des probabilités.

Pour pallier ce problème, Jiang et Allan vont remplacer la probabilité de clic par une prédiction de cette probabilité, et c'est là que les statistiques et le machine learning rentrent en jeu.

Prédire la probabilité de clic

Prédire la probabilité de clic (qu'on va souvent appeler CTR chez les SEO) est un problème qui a surtout été étudié pour les annonces sponsorisées (voir par exemple l'article [3]). Ici le contexte n'est pas le même, et la prédiction se fait avec des caractéristiques spécifiques, que nous allons maintenant voir, en même temps que la corrélation avec le taux de clic.

C'est l'information qui va intéresser les lecteurs SEO de cet article, que nous vous délivrons maintenant, et qui va vous permettre de créer des snippets qui seront plus cliqués que d'habitude !

- **Caractéristiques indépendantes de la requête tapée par l'utilisateur.**
Le premier ensemble de caractéristiques étudié par Jiang et Allan est celui des caractéristiques génériques du snippet comme sa longueur, la longueur de l'URL, etc. Ces critères ne dépendant pas de la requête, ils sont généralement faiblement corrélés à la probabilité de clic, que la page soit pertinente ou pas. Cependant, certaines corrélations restent informatives même si faibles. Par exemple, les pages qui ont la plus grande proportion de contenu réel (c'est-à-dire de contenu entre les balises <p>) sont les plus cliquées. Techniquement cela veut dire que si vous avez trop de menus, footer, etc. sur vos pages, cela se retranscrit dans le taux de clic au niveau du moteur. Une autre corrélation intéressante est négative : les pages pertinentes ont un taux de clic qui est inversement proportionnel à la longueur du titre. Plus ce dernier est long moins les pages sont cliquées.

- **Caractéristiques dépendantes de la requête tapée par l'utilisateur.** Le second ensemble étudié est celui des critères dépendants de la requête tapée. On y trouve des éléments très standard comme par exemple le nombre ou la proportion de termes de la requête dans le titre, snippet, l'URL voire dans certaines balises (on pense par exemple au H1 ou H2). On y trouve également des critères plus complexes comme par exemple le nombre de fenêtres de 10 mots consécutifs qui contiennent les termes de la requête, ou encore comme la vraisemblance de la requête, qui est une forme de vraisemblance d'indépendance, c'est-à-dire un score proportionnel à la probabilité qu'un humain lambda tape cette requête. En terme de corrélation, les plus fortes sont liées aux proportions de présence des termes de la requête dans le titre, l'URL puis dans le snippet (les éléments sont dans l'ordre décroissant de corrélation). En résumé, si vous êtes SEO, sans surprise la présence de la requête dans URL, titre et snippet augmente les chances que les utilisateurs du moteur cliquent sur vos liens. Par ailleurs, la corrélation la plus forte est avec le nombre de fenêtres de 10 mots qui contiennent la requête (voir précédemment), mais c'est sans surprise puisque plus ce nombre est grand plus la page a des chances d'être pertinente et donc plus elle va être cliquée de part sa pertinence d'abord.
- **Caractéristiques dépendantes de la session de l'utilisateur.** Il s'agit des caractéristiques portant sur le comportement de l'utilisateur d'une requête à l'autre : a-t-il reformulé la requête précédente en rajoutant, en enlevant et en gardant certains termes ? Si oui, dans quelle proportion ? Les statistiques sont assez claires, la reformulation d'une requête pour l'affiner va permettre d'améliorer la probabilité de cliquer sur la page. Quelle est l'impact de cette corrélation pour les référenceurs ? Il est simple : si vous voulez augmenter la probabilité de clic sur vos résultats dans les SERP, vous devez travailler les requêtes reformulant votre requête principale (« voiture » -> « voiture familiale », etc.).
- **Caractéristiques situationnelles.** Les critères de situation sont ceux qui sont indépendants des requêtes et des résultats, ils dépendent littéralement de la situation de l'utilisateur. On retrouve ici le nombre de clics effectués depuis le début de la session, le nombre de recherches effectuées, le temps passé entre deux clics, etc. L'idée de ces critères est qu'à un certain moment, l'utilisateur va se fatiguer naturellement de l'action de recherche et va cesser de cliquer car il va cesser de chercher, ou en tout cas cesser de chercher de manière détaillée (et donc le nombre de clics va baisser drastiquement). Ces caractéristiques ne sont pas corrélées à la probabilité de clic pour les pages pertinentes, mais elles le sont pour les pages non pertinentes. C'est une information intéressante mais naturelle : l'éventuelle lassitude de recherche a un impact sur la probabilité de cliquer sur les pages moins pertinentes, mais pas sur la probabilité de cliquer sur des pages

pertinentes, qu'il faudra de toute façon voir pour satisfaire son besoin informationnel.

Toutes ces caractéristiques et corrélations sont agrégées au sein d'une formule qui va donner une estimation de la probabilité de clic. Nous renvoyons le lecteur intéressé par les formules exactes à l'article [1].

La vraie vie : la prédiction ce n'est pas facile

De tout ce que nous avons vu précédemment, nous utilisons les probabilités de clic pour faire le classement, et ce qui apparaît implicitement est que la prédiction de ces probabilités se fait à l'aide des caractéristiques les plus corrélées. Or, réaliser cette prédiction à partir des caractéristiques corrélées est extrêmement difficile. Pour pallier ce problème, Jiang et Allan ont utilisé un algorithme de *machine learning* qui utilise des réseaux de neurones pour classer des éléments de manière optimale (voir l'article [4] par Christopher Burges, chercheur chez Microsoft Research).

A l'aide de cet algorithme, ils ont pu estimer les probabilités de clic pour ensuite modifier la position des pages proposées dans un classement déjà existant. Leur outil a de bonnes performances puisque dans le pire cas, l'erreur qu'il commet est de 13% (en absolu, c'est-à-dire que la probabilité réelle est plus ou moins 13% de la prédiction réalisée). L'erreur dans le meilleur cas est de 8%.

Par ailleurs, un autre problème qu'ils ont dû prendre en compte est celui des nouvelles requêtes. En effet, pour prédire les probabilités de clic, il faut posséder un historique concernant les sessions de recherche pour cette requête, ce qui n'est pas toujours le cas. Plusieurs solutions sont évoquées pour ce problème, la plus simple (et moins efficace) étant de regrouper les requêtes par thématique.

Enfin, Jiang et Allan ont testé leur méthodologie en reclassant la première page des SERP pour quelques centaines de requêtes. Les résultats montrent une amélioration des taux de clics sur les pages les plus pertinentes, et donc une diminution des sauts de ces mêmes pages.

Conclusion

Ce travail est extrêmement intéressant, car il formalise pour la première fois l'utilisation d'une notion supplémentaire à la popularité (le PageRank) et la pertinence : l'attractivité des résultats présentés par le moteur.

Jiang et Allan ont même une observation assez amusante dans leurs résultats d'expérience : l'amélioration sur les clics est de 10 à 20%, mais la pertinence des pages proposées décroît de 2,1% environ. En utilisant cette méthode de classement, on augmente la satisfaction des utilisateurs en baissant pourtant la qualité moyenne des pages proposées ! Un comble...

Références

[1] Jiepu Jiang and James Allan. 2016. Reducing Click and Skip Errors in Search Result Ranking. In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM '16). ACM, New York, NY, USA, 183-192.

https://people.cs.umass.edu/~jpjiang/papers/wsdm16_rel_click.pdf

[2] Emine Yilmaz, Milad Shokouhi, Nick Craswell, and Stephen Robertson. 2010. Expected browsing utility for web search evaluation. In Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM '10). ACM, New York, NY, USA, 1561-1564.

[3] Ramakrishnan Srikant, Sugato Basu, Ni Wang, and Daryl Pregibon. 2010. User browsing models: relevance versus examination. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '10). ACM, New York, NY, USA, 223-232.

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.361.2986&rep=rep1&type=pdf>

[4] BURGES, Christopher JC. From ranknet to lambdarank to lambdamart: An overview. Learning, 2010, vol. 11, p. 23-581.

<https://www.microsoft.com/en-us/research/publication/from-ranknet-to-lambdarank-to-lambdamart-an-overview/>



Guillaume Peyronnet est gérant de Nalrem Médias. **Sylvain Peyronnet** est cofondateur et responsable des ix-labs, un laboratoire de recherche privé. Ensemble, ils font des formations, pour en savoir plus : <http://www.peyronnet.eu/blog/>