Evaluation des systèmes de recherche d'information interactifs : exemple du Bouclage de Pertinence implicite par suivi oculaire

Philippe Mulhem, Francis Jambon, Lucas Albarede

▶ **To cite this version:**

Philippe Mulhem, Francis Jambon, Lucas Albarede. Evaluation des systèmes de recherche d'information interactifs : exemple du Bouclage de Pertinence implicite par suivi oculaire. Journée Accès Interactif à l'Information, Apr 2022, Paris, France. pp.14-16. hal-03634132

HAL Id: hal-03634132
https://hal.science/hal-03634132

Submitted on 7 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluation des systèmes de recherche d'information interactifs : exemple du Bouclage de Pertinence implicite par suivi oculaire

Philippe Mulhem, Francis Jambon, Lucas Albarede
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
{Philippe.Mulhem, Francis.Jambon}@imag.fr,
lucas.albarede@etu.univ-grenoble-alpes.fr

1 Introduction

Un système de recherche d'information (SRI) est par définition un système interactif qui vise à fournir à un utilisateur les documents pertinents pour son besoin d'information. L'évaluation des SRI est un élément fondamental de ce domaine de recherche. Historiquement, l'aspect interactif a été peu intégré dans l'évaluation des SRI : il rentre en conflit avec le fait que l'on souhaite évaluer les systèmes de manière reproductible et "en grand", c'est-à-dire avec de nombreuses requêtes portant sur de nombreux documents, et qu'il n'est souvent pas possible, d'un simple point de vue logistique, de réaliser cela avec les humains. Nous proposons une approche basée sur la simulation d'un ensemble d'utilisateurs afin de se confronter à ce problème, en prenant comme exemple le cas du bouclage de pertinence avec du suivi oculaire, pour la recherche de pages web avec utilisation de snippets.

2 Exemple du bouclage de pertinence implicite par suivi oculaire

Le bouclage de pertinence implicite utilisant le suivi oculaire s'appuie sur l'hypothèse qu'il est possible de déterminer les centres d'intérêts implicites d'un utilisateur, en étudiant la façon dont celui-ci lit les termes d'une page de résultats (Sungeelee *et al.*, 2020). L'état de l'art sur ce sujet étant restreint, nous avons créé notre propre modèle de bouclage de pertinence qui intègre des caractéristiques obtenues à partir du suivi oculaire d'un utilisateur. Ce modèle, appelé modèle Bo1ET (Albarede, 2019), est une extension du modèle Bo1 (Amati & Van Rijsbergen, 2002) de pseudo-bouclage de pertinence, que nous avons adapté pour la recherche sur le Web (SERP), et intégrant des caractéristiques "classiques" (Holmqvist *et al.*, 2011) du suivi oculaire, comme le temps de lecture moyen d'un mot. Le paramétrage puis l'évaluation "en grand" d'un tel système, tout en intégrant des éléments provenant d'une interaction avec l'utilisateur reproductible, nécessiterait d'avoir recours à plusieurs centaines d'utilisateurs (afin de minimiser la variabilité inter-utilisateurs), sur l'ensemble d'une collection (afin d'être comparable avec d'autres travaux), et dans des locaux dédiés (afin de pouvoir utiliser le dispositif de suivi oculaire), ce qui n'est pas réalisable en pratique. C'est pourquoi, nous nous sommes tournés vers l'utilisation de la simulation du comportement des utilisateurs.

3 Cadre général de l'évaluation par simulation

Quand on parle de simulation de suivi oculaire, la question de quoi simuler, et pour quel objectif, se pose. Dans notre cas, nous avons choisi de simuler non pas directement les parcours oculaires, mais les caractéristiques de haut niveau : notre argument principal de ce choix est que les simulations des parcours oculaires sont d'une très grande complexité (Yarbus, 1967), et pas nécessaire dans notre cas. Notre système reposant sur des caractéristiques générées à partir des parcours oculaires,

nous choisissons de privilégier la simulation au niveau de ces caractéristiques : dans ce cas, nous ne simulons que les éléments directement liés à notre problématique. Le désavantage d'une telle solution est qu'il n'est possible d'évaluer que les systèmes interactifs utilisant exclusivement ces caractéristiques. Le choix de ces caractéristiques est donc d'un intérêt primordial pour l'acceptation de nos propositions. De manière plus formelle, nous appelons :

— caractéristiques hors contexte applicatif, F_{hc}, les suivis oculaires (saccades et fixations). Ces éléments ont classiquement acquis par les eye-trackers quelque soit le contexte applicatif considéré ;

— caractéristiques en contexte applicatif, F_{ec}, les caractéristiques construites à partir des F_{ec} qui décrivent des éléments du domaine applicatif. Dans notre cas, ces caractéristiques sont celles utilisées par un SRI interactif, par exemple, le temps de lecture d'un mot pertinent, le temps de lecture d'une snippet, etc.

Le passage des F_{hc} au F_{ec} est donc fondamental et doit être défini avec précision.

4 Adaptation du cadre général à l'évaluation de la RF

De manière à pouvoir comparer notre approche à d'autres travaux, et à la maintenir dans le paradigme de Cranfield, nous utilisons les documents du corpus, les requêtes et les jugements de pertinence issus de collections de test classiques. Nous avons étendu ces données pour l'évaluation interactive en y ajoutant des indications sur la pertinence des mots présentés dans les SERP, et la pertinence des snippets (déduite directement de celle des documents). Dans notre exemple, chaque caractéristique de contexte applicatif F_{ec} (ou feature) est associée à une probabilité de valeur pour chaque mot de chaque snippet présenté. Cette feature est calculée à partir des caractéristiques associées au mot qui dépendent de la pertinence du mot (3 valeurs de pertinences : pertinent/non-pertinent/topical), et des caractéristiques associées à la snippet (2 valeurs de pertinences : pertinent/non-pertinent). Notons que les valeurs de pertinences sont connues du générateur de features, qui simule l'humain, mais pas du SRI. Nous restons donc cohérent avec le fait que le SRI réagit uniquement aux actions simulées, actions qui elles dépendent de la pertinence des éléments présentés à l'interface, mais le SRI n'a pas connaissance des valeurs de pertinence. Dans notre exemple, nous avons défini par un jugement d'experts la pertinence des mots des snippets, en se basant sur les requêtes initiales. La pertinence d'un snippet est déduite de la pertinence des documents définie dans la collection de test (données existantes dans les collections de test actuelles)(Mizzaro, 1997).

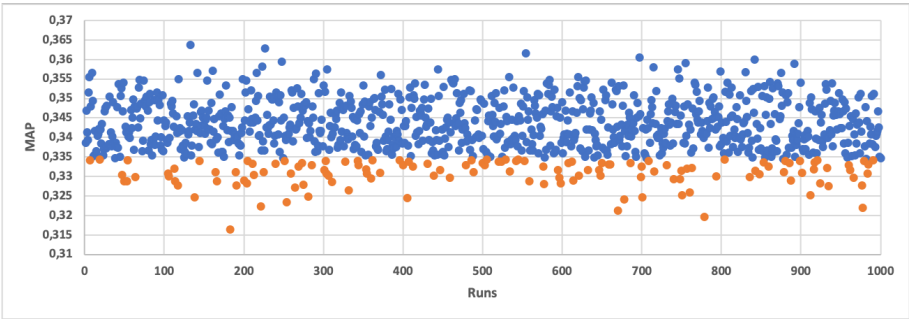

FIGURE 1 – MAP de Bo1EG pour 1000 utilisateurs simulés ; axe des X : numéro d'utilisateur ; axe des Y : MAP de Bo1EG comparée à Bo1mod (bleu : Bo1EG est supérieur à Bo1mod ; rouge : Bo1EG est inférieur à Bo1mod).

5 Mise en œuvre

Une première étude a été effectuée sur 8 requêtes sur les corpus TREC-5 et TREC-6, avec les requêtes de TREC-6 ad-hoc (Voorhees & Harman, 2000). Dans un premier temps, la feature utilisée a été mesurée sur un ensemble restreint d'utilisateurs de manière à en déterminer sa gaussienne. Elle a ensuite été simulée à partir de cette dernière. La figure 1 présente les résultats sur 1000 simulations qui montrent un avantage (en MAP) que Bo1ET améliore dans 867 cas (points bleus) les résultats de Bo1 classique, et dégrade dans 133 cas (points rouges).

6 Conclusion et Perspectives

Ces travaux préliminaires montrent qu'il est envisageable, tout en restant dans le cadre du paradigme de Cranfield, d'évaluer un système de recherche d'information interactif via la simulation des utilisateurs. Nous travaillons actuellement à généraliser et outiller cette approche afin d'étendre son domaine de validité et d'utilisation.

7 Remerciements

Ces travaux ont été en partie financés par le Laboratoire d'Informatique de Grenoble (UMR 5217) via les projets "Emergence" GELATI et IIRSuts.

Références

ALBAREDE L. (2019). *Relevance Feedback Enhanced with Eye Movements Analysis*. Rapport interne, Université Grenoble Alpes, Master of Science in Informtics at Grenoble.

AMATI G. & VAN RIJSBERGEN C. J. (2002). Probabilistic models of information retrieval based on measuring the divergence from randomness. *ACM Trans. Inf. Syst.*, **20**(4), 357–389.

HOLMQVIST K., NYSTRÖM M., ANDERSSON R., DEWHURST R., HALSZKA J. & VAN DE WEIJER J. (2011). *Eye Tracking : A Comprehensive Guide to Methods and Measures*. United Kingdom : Oxford University Press.

MIZZARO S. (1997). Relevance : The whole story. *Journal of the American Society for Information Science*, (48), 810–832.

SUNGEELEE V., JAMBON F. & MULHEM P. (2020). Proof of Concept and Evaluation of Eye Gaze Enhanced Relevance Feedback in Ecological Context. In *Proceedings of the Joint Conference of the Information Retrieval Communities in Europe (CIRCLE 2020), Samatan, Gers, France, July 6-9, 2020*, volume 2621 de *CEUR Workshop Proceedings* : CEUR-WS.org.

VOORHEES E. M. & HARMAN D. (2000). Overview of the sixth text retrieval conference (trec-6). *Inf. Process. Manage.*, **36**(1), 3–35. DOI : 10.1016/S0306-4573(99)00043-6.

YARBUS A. L. (1967). *Eye Movements and Vision*. Plenum. New York.