

# Conception d'un modèle généraliste pour l'évaluation d'un test A/B

Emmanuelle Claeys\*, Pierre Gançarski \*  
Myriam Maumy-Bertrand\*\* Hubert Wassner\*\*\*

\*ICube – Université de Strasbourg – 67412 – Illkirch – France  
{claeys, gançarski}@unistra.fr

\*\*IRMA – Université de Strasbourg – 67084 – Strasbourg – France  
mmaumy@math.unistra.fr

\*\*\*Entreprise AB Tasty – 3 Impasse de la Planchette – 75003 – Paris – France  
hubert@abtasty.com

## 1 La problématique d'un test A/B

Un test *A/B* (ou *A/B testing*<sup>1</sup>), consiste à créer plusieurs versions d'un même objet, dans le but de vérifier une hypothèse marketing. Par exemple, il s'agit de tester auprès d'un échantillon d'internautes, nommés visiteurs, durant une période donnée, différentes variations (généralement deux) d'une même page web (nommées variation *A* et variation *B*) proposées par un e-commerçant. Le but est d'identifier celle qui donne les meilleurs résultats par rapport à un objectif fixé par l'e-commerçant en amont du test. Une contrainte commune, lors d'un test *A/B*, est que lorsqu'un visiteur est affecté à une variation, elle lui est affectée pour toute la durée de sa visite, jusqu'à la fin du test. Il est donc impossible de connaître de façon sûre le comportement qu'il aurait eu sur l'autre variation. De plus, les visiteurs testés sur la variation *A* étant différents de ceux testés sur la variation *B*, il est nécessaire de définir un moyen de comparer leurs comportements respectifs. Le site étant généralement en production, il est intéressant de garder un compromis entre l'exploration (évaluer les différentes variations) et l'exploitation du test (favoriser la variation avec les meilleurs résultats). Sachant qu'il est impossible de modifier la variation appliquée à un visiteur, il est nécessaire d'affecter au mieux à un nouveau visiteur, la version *A* ou *B* du test.

## 2 Modélisation et approche multi-contextes

Plusieurs verrous majeurs sont à lever : le modèle décisionnel doit prendre en compte la performance et la pertinence de ses choix. Il devra adapter ses choix en fonction du type de visiteurs et du contexte du test. Le modèle devra être également capable d'apprendre de ses succès/erreurs et de s'auto-corriger.

---

1. Le terme *A/B* étant un terme générique pour qualifier tout type de tests

## Conception d'un modèle généraliste pour l'évaluation d'un test A/B

Enfin, ce modèle doit être suffisamment générique pour pouvoir être utilisé dans différentes configurations : type de e-commerce, type de tests réalisés, période de test, etc.

Actuellement, lorsqu'un e-commerçant souhaite réaliser sur le site un test A/B, il fait appel à un expert chargé d'examiner son site. Ce dernier étudie le trafic, les types de transactions réalisées, et lui demande de choisir ce qu'il souhaite comparer. Avec l'aide de l'expert, l'e-commerçant choisit de modifier un élément de sa page puis débute le test.

L'objectif de nos travaux est de proposer une méthodologie en quatre étapes permettant d'automatiser au mieux ce processus de mise en oeuvre d'un test A/B<sup>2</sup>.

Dans une première étape, l'e-commerçant va être classé automatiquement dans une catégorie de commerces. Classifier l'e-commerçant permet de partir avec un *a priori* lors de l'analyse. Une fois que l'e-commerçant est classé, dans une deuxième étape, la variation que l'e-commerçant souhaite appliquer par rapport à sa page originale doit être étudiée. De même que pour la classification des e-commerçants, une classification des tests permettra, en fonction des objectifs, de mieux anticiper l'impact d'une variation et la durée qu'il faudrait allouer au test.

Dans une troisième étape, des groupes de visiteurs vont être créés à partir du trafic sur le site. Les visiteurs testés sur la variation A étant différents de ceux testés sur la variation B, nous proposons de trouver des visiteurs aux caractéristiques similaires et de comparer leurs comportements après avoir été soumis au test. De plus, si une variation peut être optimale pour la majorité des visiteurs, il est plus intéressant pour un e-commerçant d'observer pour quelle catégorie de visiteurs cette variation est optimale/sous-optimale. Anticiper le comportement de nos visiteurs nous permet d'explorer la variation A avant même d'avoir commencé le test (et ainsi anticiper le gain de la variation A). Le test identifie alors plus rapidement la variation "gagnante" pour chaque clusters de visiteurs.

Lors de la dernière étape, correspondant au test lui-même, pour un cluster de visiteurs choisi, la régression permettra de comparer les résultats de la variation B avec l'originale A. Elle est réalisée pour chaque segment (cluster) de visiteurs ce qui permettra de détecter éventuellement une différence de résultats importante entre A et B pour un segment de visiteurs particuliers ou au contraire, de ne pas constater de différences significatives et de continuer l'exploration. Lorsque l'écart moyen entre la courbe de la variation A et de la variation B dépasse un certain seuil donné, le modèle décidera alors d'allouer exclusivement A ou B.

### 3 Conclusion

Dans nos expériences, nous avons réalisé une classification ascendante hiérarchique pour évaluer la similarité entre le trafic des e-commerces. Nous avons ainsi pu construire des groupes d'e-commerçants similaires, dans le but d'identifier différents comportements de visiteurs. Nous avons également extrait les mots clefs associés aux pages qu'a parcouru un visiteur sur le site avant d'arriver sur la page test. Cette expérience a pour but d'enrichir les informations d'un visiteur pour améliorer la qualité de nos clusters. Pour conclure, l'objectif de cette méthode, pour l'amélioration d'un test A/B par rapport aux techniques existantes, est de proposer une allocation de variation basé sur la classification des e-commerçants, de celle des tests et enfin de celle des visiteurs.

---

2. Il appartiendra au e-commerçant, à la suite du test, de faire un affichage personnalisé définitif à ses visiteurs, si le test est significatif, pour un ou plusieurs segments