

Prévisions d'audiences et optimisation de plannings de publicités internet

Antoine Jeanjean, Bruno Martin

Bouygues e-lab, Paris, {ajejanjean, bmartin}@bouygues.com

1 Introduction

Nous détaillons ici les algorithmes utilisés dans un outil d'aide à la décision créé pour le groupe TF1. Le problème consiste ici à prévoir les audiences de chaque tag de publicités d'une collection de sites internet. En utilisant ces données de prévisions dans un moteur de simulation, on peut déduire le "reste à vendre" de chaque offre publicitaire, soit le stock non vendu dans les prochaines semaines. Ce stock représente le nombre d'affichages de pages disponibles pour y intégrer des bannières publicitaires. L'algorithme de prédiction d'audience est capable de prévoir de manière précise l'audience d'une cinquantaine de sites internet (TF1.fr, Wat.tv, Eurosport.fr, over-blog.com, etc.) composés chacun d'une collection de tags de publicités arborescents. Cet algorithme est utilisé depuis septembre 2009 pour prévoir les audiences quotidiennes et anticiper les encombrements de planning. Les tests, réalisés sur les données réelles de 5 sites, montrent que le taux d'erreur moyen reste inférieur à 10 %.

2 Prévisions d'audiences internet

Le nombre de papiers traitant de prévisions d'audiences (internet ou TV) est limité. Yahoo Research a publié des travaux qui ne détaillent que très peu les algorithmes sous-jacents en se focalisant davantage sur la probabilité de taux de clics [1]. Les publications sur les prévisions d'audiences télévisuelles traitent plutôt d'analyse de panels [2,3]. Au coeur de serveur de distribution de publicités, les prévisions sont bien souvent des extrapolations linéaires du passé. Il n'existe d'ailleurs aucun logiciel du commerce pour les audiences web. Le problème reste très dépendant de la manière de commercialiser son offre publicitaire et de gérer ses stocks et de nombreux facteurs, trouvant leurs racines dans le coeur même de l'activité, influencent l'audience. Nous présentons ici une heuristique arborescente appelée récursivement sur chaque noeud de l'arbre et optimisée pour prévoir, en quelques minutes, plusieurs semaines d'audiences à partir d'une année d'historique. Les sites ont une liste de rubriques ayant eux-mêmes des sous-rubriques. La profondeur de cet arbre peut atteindre une vingtaine de niveaux. Pour chaque sous-rubrique, on connaît l'audience moyenne de ses tags, utile pour déterminer la méthode de prédiction qui va être sélectionnée. Si cette moyenne est plus petite qu'un certain seuil S ou si on ne dispose pas d'un historique de données suffisant, l'audience de la sous-rubrique n'est pas prédite à partir de l'algorithme de prédiction mais déduite de ces sous-rubriques "soeurs" et de sa rubrique "mère". Si l'audience respecte les critères de prédiction, on procède tout d'abord à un gommage de l'influence des événements passés de cette sous-rubrique. Sur cet historique purifié, on est en mesure de lancer un apprentissage sur une liste d'indicatrices variées. Ces indicatrices, connues dans le passé comme dans le futur, vont

permettre d'associer des impacts à différents critères : indicatrice des vacances d'été, des lundis d'hiver, des ponts de trois jours, la tendance linéaire, la tendance sinus, etc. Pour chaque sous-rubrique, on teste si l'ajout de la nouvelle indicatrice ne vient pas perturber l'apprentissage des autres indicatrices. Puis, on est en mesure de calculer les coefficients de chaque facteur à l'aide d'une régression linéaire par moindres carrés multiples. On en déduit ensuite l'audience globale en utilisant les indicatrices futures ainsi que les impacts des éventuels événements impactant les jours à venir, avec ou sans traînée d'audience. Pour toutes les rubriques qui n'ont pas été prédites faute d'informations ou de volumes suffisants, on déduit leur audience par un apprentissage de leur représentativité face aux audiences des sous-rubriques du même noeud. Cette méthode, à la fois simple, robuste et efficace, permet de calculer rapidement l'audience prévisionnelle d'un grand nombre de sites tout en captant les tendances, les impacts brutaux et les influences locales du site.

3 Simulation de l'outil de diffusion

La simulation de la diffusion des publicités internet permet de répondre à une question essentielle lors de la commercialisation d'une campagne : "Est-il possible d'accepter cette campagne en conservant une probabilité raisonnable de diffuser l'ensemble des campagnes vendues?". Une bonne prévision des audiences de ses sites internet n'est pas suffisante pour prendre de bonnes décisions commerciales. En effet, il ne suffit pas toujours d'avoir un nombre d'affichages suffisant pour que les campagnes vendues aux annonceurs soient diffusées correctement. De plus, la simulation permet de suivre la diffusion des campagnes et de déterminer les périodes pendant lesquelles il est opportun d'envoyer une newsletter ou de lancer un jeu pour faire augmenter les audiences et permettre la diffusion de toutes les campagnes. Ce moteur de simulation doit respecter l'ensemble des contraintes (ciblage sur un sous-ensemble de sites, ciblage géographique, cookies de répétition, etc.). La difficulté algorithmique d'un outil de simulation vient presque toujours des exigences de performances. De plus, il est important de savoir répondre à l'annonceur en "temps réel", c'est à dire dans notre cas en quelques secondes. Simuler tous les affichages sur toutes les pages sur plusieurs semaines est dans ce cas hors de question. Il faut calculer analytiquement la répartition des publicités sur une page en prenant en compte les priorités des campagnes ainsi que les limites par visiteur unique imposées par les campagnes ou par les formats. Connaissant les prévisions d'audiences et la simulation de la diffusion, l'étape suivante consiste à améliorer la planification de ces publicités en utilisant un moteur de règles. Ces recommandations permettent d'influencer le moteur de diffusion, en lui indiquant des changements de dates ou de priorités de campagnes par exemple, afin d'optimiser la consommation du stock d'affichages disponibles.

Références

1. X. Wang, A. Broder, M. Fontoura, V. Josifovski. A Search-based Method for Forecasting Ad Impression in Contextual Advertising. In *18th International WWW Conference*, April 2009.
2. P. Cattin, R. Festa, A. Le Diberder, A Model for Forecasting the Audience of TV Programs. In : *Worldwide Electronic and Broadcasting Audience Research Symposium*, Ū Vol. 1. Pages 513-524.
3. Weber, René. Methods to Forecast Television Viewing Patterns for Target Audiences. In : *Communication Research in Europe and Abroad Ū Challenges of the First Decade*. Berlin : DeGruyter.