Bandits manchots avec échantillonnage de Thompson pour des recommandations multiples suivant un modèle fondé sur les positions

Camille-Sovanneary Gauthier * Romaric Gaudel † Elisa Fromont ‡

Résumé

Les systèmes de recommandation en ligne ont pour but de proposer les produits les plus intéressants aux positions adéquates sur une page internet. Nous présentons un nouvel algorithme, PB-MHB, permettant de faire des recommandations multiples en ligne en suivant un modèle fondé sur les positions. Cet algorithme s'appuie sur le principe des bandits manchots et utilise un échantillonnage de Thompson couplé avec un algorithme de Metropolis-Hastings pour tirer les paramètres des lois probabilistes utilisées, ce qui n'avait jamais été fait dans le contexte d'un modèle basé positions. Notre méthode ne nécessite pas d'avoir en paramètre les probabilités de vue des utilisateurs sur chaque position de la page Web, comme cela est usuellement le cas pour les algorithmes répondant à ce type d'interaction. Celles-ci sont d'ailleurs en pratique difficile à obtenir *a priori*. Les expériences faites sur des données simulées et sur des données issues de bases de données réelles (KDD-CUP2012 et Yandex) montrent que notre méthode, avec moins d'information, fournit de meilleurs recommandations que l'état de l'art.

References

Shipra Agrawal and Navin Goyal. 2013. Thompson Sampling for Contextual Bandits with Linear Payoffs. In *proc. of the 30th Int. Conf. on Machine Learning (ICML '13)*.

Shipra Agrawal and Navin Goyal. 2017. Near-Optimal Regret Bounds for Thompson Sampling. *Jour. of the ACM, JACM* 64, 5, Article 30 2017,

Maarten de Rijke Aleksandr Chuklin, Ilya Markov. 2015. Click Models for Web Search. In *Click Models for Web Search*.

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. 2002. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning* 47, 2 (01 May 2002), 235–256.

Olivier Chapelle and Lihong Li. 2011. An Empirical Evaluation of Thompson Sampling. In *Advances in Neural Information Processing Systems, NIPS'11*.

Wei Chen, Yajun Wang, and Yang Yuan. 2013. Combinatorial multi-armed bandit: General framework and applications. In *proc. of the Int. Conf. on Machine Learning (ICML '13)*.

Wang Chi Cheung, Vincent Tan, and Zixin Zhong. 2019. A Thompson Sampling Algorithm for Cascading Bandits. In *proc. of the 22nd Int. Conf. on Artificial Intelligence and Statistics, AISTATS'19*.

Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. 2015. *Click Models for Web Search*.

Richard Combes, Stefan Magureanu, Alexandre Proutiere, and Cyrille Laroche. 2015. Learning to Rank: Regret Lower Bounds and Efficient Algorithms. In *proc. of the ACM Int. Conf. on Measurement and Modeling of Computer Systems*.

Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An Experimental Comparison of Click Position-bias Models. In *proc. of the Int. Conf. on Web Search and Data Mining (WSDM '08)*.

Bianca Dumitrascu, Karen Feng, and Barbara Engelhardt. 2018. PG-TS: Improved Thompson Sampling for Logistic Contextual Bandits. In *Advances in Neural Information Processing Systems 31, NIPS'18*.

Sumeet Katariya, Branislav Kveton, Csaba Szepesvari,

*Louis Vuitton - Univ. Rennes
†Univ. Rennes, ENSAI, CNRS, CREST - UMR 919
‡Univ. Rennes, IRISA/INRIA rba, IUF

Claire Vernade, and Zheng Wen. 2017a. Stochastic Rank-1 Bandits. In *proc. of the Int. Conf. on Artificial Intelligence and Statistics, AISTATS'17*.

Sumeet Katariya, Branislav Kveton, Csaba Szepesvári, Claire Vernade, and Zheng Wen. 2017b.Bernoulli Rank-1 Bandits for Click Feedback. In *proc. of the International Joint Conference on Artificial Intelligence, IJCAI'17*.

Jaya Kawale, Hung H Bui, Branislav Kveton, Long Tran-Thanh, and Sanjay Chawla. 2015. Efficient Thompson Sampling for Online Matrix Factorization recommandation. In *Advances in Neural Information Processing Systems 28, NIPS'15*.

Junpei Komiyama, Junya Honda, and Hiroshi Nakagawa. 2015. Optimal Regret Analysis of Thompson Sampling in Stochastic Multi-armed Bandit Problem with Multiple Plays. In *proc. of the 32nd Int. Conf. on Machine Learning (ICML '15)*.

Junpei Komiyama, Junya Honda, and Akiko Takeda. 2017. Position-based Multiple-play Bandit Problem with Unknown Position Bias. In *Advances in Neural Information Processing Systems 30, NIPS'17*.

Branislav Kveton, Chang Li, Tor Lattimore, Ilya Markov, Maarten de Rijke, Csaba Szepesvári, and Masrour Zoghi. 2018. BubbleRank: Safe Online Learning to Rerank. *CoRR* abs/1806.05819 (2018).

Branislav Kveton, Csaba Szepesvári, Zheng Wen, and Azin Ashkan. 2015a. Cascading Bandits: Learning to Rank in the Cascade Model. In *proc. of the Int. Conf. on Machine Learning (ICML '15)*.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvári. 2015b. Combinatorial Cascading Bandits. In *Advances in Neural Information Processing Systems 28 (NIPS '15)*.

Paul Lagrée, Claire Vernade, and Olivier Cappé. 2016. Multiple-play Bandits in the Position-based Model. In *Advances in Neural Information Processing Systems 30 (NIPS '16)*.

Tor Lattimore, Branislav Kveton, Shuai Li, and Csaba Szepesvari. 2018. TopRank: A practical algorithm for online stochastic ranking. In *NIPS'18*.

Shuai Li, Baoxiang Wang, Shengyu Zhang, and Wei Chen. 2016. Contextual Combinatorial Cascading Bandits. In *proc. of Int. Conf. on Machine Learning, ICML'16*.

Eric Mazumdar, Aldo Pacchiano, Yi-An Ma, Peter L. Bartlett, and Michael I. Jordan. 2020. On Thompson Sampling with Langevin Algorithms. *CoRR* abs/2002.10002 (2020).

Radford M. Neal. 1993. *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Technical Report CRG-TR-93-1. University of Zurich, Department of Informatics.

Matthew Richardson, Ewa Dominowska, and Robert Ragno. 2007. Predicting Clicks: Estimating the Click-Through Rate for New Ads. In *proc. of the 16th International World Wide Web Conference (WWW '07)*.

Carlos Riquelme, George Tucker, and Jasper Snoek. [n.d.]. Deep Bayesian Bandits Showdown: An Empirical Comparison of Bayesian Deep Networks for Thompson Sampling. In *proc. of the Int. Conf. on Learning Representations, ICLR'18*.

William R. Thompson. 1933. On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika* 25, 3/4 (1933), 285–294.

John von Neumann. 1951. Various Techniques Used in Connection with Random Digits. In *Monte Carlo Method*, 12, 36–38.

Yandex. 2013. Yandex personalized web search challenge.

Masrour Zoghi, Tomas Tunys, Mohammad Ghavamzadeh, Branislav Kveton, Csaba Szepesvari, and Zheng Wen. 2017. Online Learning to Rank in Stochastic Click Models. In *proc. of the 34th Int. Conf. on Machine Learning, ICML'17*

Shi Zong, Hao Ni, Kenny Sung, Nan Rosemary Ke, Zheng Wen, and Branislav Kveton. 2016. Cascading Bandits for Large-scale recommandation Problems. In *proc. of the 32nd Conference on Uncertainty in Artificial Intelligence (UAI '16)*.