Plan pracy na 4 marca

Hubert Drażkowski

04.03.2022

1 Plan pracy

1. An introduction (no mathematical notation, brevity) [2] [16]
   - A brief description of what is the problem of multi armed bandits and online learning with a special emphasis on adding the context.
   - Arguments for why this topic is interesting from mathematical, informatics and real life application side, again stressing the contextual variant. [16] [2] [8]
   - Stating the distinction between reinforecement learning, game theory, decision theory and multi armed bandits.
   - Stating what is the main problem of the thesis and main challange.
   - Showing the novelty of the work in the contrast to what was written already and showing the scientific impact it might have.
   - A brief description on what is covered by subsequent paragraphs, how they are linked to each other.
2. Part I Notation and an introduction to the topic
   (a) Introducing standard probability space, canonical bandit model, useful concentrations of measures. Chebyshev inequality, Bernstein inequality, subgaussian random variables, Cramér Chernoff methods [2] [22]
   (b) Introducing notation and main general definitions. What is a reward, an action, a history, an enviroment, an arm, a measure of quality of an algorithm, a regret, a context in different forms, a competing class. Decomposition of a regret. [5]
   (c) Possible enviroments and its applications, a historical note.
   (d) Lower bound. Mathematical assumptions discussed.
3. Part II Classical MAB (for building intuitions and historical notes, base for the contextual versions)
   (a) ϵ greedy algorithms and ETC algorithm
   (b) UCB algorithms family
   (c) Thompson Sampling [12] [13] [14]
4. Part III Linear Contextual MAB
   (a) A context setting [5]
   (b) A multivariate parametric context [29]
   (c) A multivariate non parametric context [30]
   (d) Lipschitz condition assumption, lipschitz bandits [18] [17] [23] [3]
   (e) Linear bandits [2] [8] [10] [24] [2]
      - least squares, confidence levels, sparsity, asymptotics, minimax lower bound [2]
   (f) Generalized linear case [31]
   (g) Side information, expert case, VC dimensions [27] [2] [6] [28]
   (h) Bayesian interpretation of contextual bandits
   (i) Algorithms
      i. Epoch greedy [33]
      ii. Methods of mixtures [2]
      iii. Reduction of EXP3 CMAB to MAB [1]

iv. EXP4 [1]
v. LinUCB, SupLinUCB, intuition, algorithm and proof of the bounds on the regret [8] [9] [10]
vi. LinREL, SupLinREL [11]
vii. PUCB [15]
viii. Thompson sampling with linear payoffs, CofineUCB [13] [14]
ix. Exploitation only algorithms [25] [26]

5. Part IV Comparison
   - Theoretical synthesis and comparison of algorithms in a synthetic way. For example a possible enviroment extensions, regrets/regret bounds etc.
   - A comparison of an algorithms on a synthetic dataset. Reduced to MAB as a banchmark, all of the before mentioned in use.
   - A comparison of algorithms on a real dataset. Reduced to MAB as a benchmark, all of the before mentioned in use.
6. Conclusions (brevity, results, should be read as one part with an introduction)
   - Complementary to the introduction, a refreshment of what was done in the paragraphs.
   - What is the answer to the posted problem.
   - What are the specific results and main conclusions of the work.
   - What are possible extensions to the work.

References

[1] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. SIAM Journal on Computing, 32(1):48–77, 2002

[2] Tor Lattimore, Csaba Szepesvári (2020). Bandit Algorithms, Cambridge University Press

[3] Aleksandrs Slivinks (2019). Introduction to Multi-Armed Bandits, Foundations and Trends in Machine Learning, Vol 12, No 1-2, 1-286.

[4] Li Zhou (2015). A survey on Contextual Multi-armed Bandits, arXive.

[5] Chih-Chun Wang, Sanjeev Kulkarani, Vincent Poor (2005). Bandit problems with side observations, IEEE Transactions on Automatic Control, 50, 338-355.

[6] Sebastien Bubeck, Nicolo Cesa – Bianchi Regret (2012). Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems, Now Publihsers

[7] Joannés Vermorel, Mehryar Mohri (2005) Multi-Armed Bandit Algotithms and Empirical Evaluation, Machine Learning : EMCL 2005, Springer, 437-448.

[8] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In Proceedings of the 19th International Conference on World Wide Web, pages 661–670. ACM, 2010.

[9] Wei Chu, Lihong Li, Lev Reyzin, and Robert E Schapire. Contextual bandits with linear payoff functions. In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, pages 208–214, 2011

[10] Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Improved algorithms for linear stochastic bandits. In Advances in Neural Information Processing Systems 24, pages 2312–2320, 2011

[11] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. The Journal of Machine Learning Research, 3:397–422, 2003

[12] WILLIAM R THOMPSON, ON THE LIKELIHOOD THAT ONE UNKNOWN PROBABILITY EXCEEDS ANOTHER IN VIEW OF THE EVIDENCE OF TWO SAMPLES, Biometrika, Volume 25, Issue 3-4, December 1933, Pages 285–294

[13] Shipra Agrawal and Navin Goyal. Further optimal regret bounds for thompson sampling. In International Conference on Artificial Intelligence and Statistics, pages 99–107, 2013b

[14] Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In Advances in Neural Information Processing Systems 24, pages 2249–2257, 2011.

[15] Rosin, C.D. Multi-armed bandits with episode context. Ann Math Artif Intell 61, 203–230 (2011).

[16] Bouneffouf, D., Rish, I., Aggarwal, C. (2020). Survey on Applications of Multi-Armed and Contextual Bandits. 2020 IEEE Congress on Evolutionary Computation (CEC).

[17] Tyler Lu, D´avid P´al, and Martin P´al. Showing Relevant Ads via Lipschitz Context Multi-Armed Bandits. In 14th Intl. Conf. on Artificial Intelligence and Statistics (AISTATS), 2010.

[18] Elad Hazan and Nimrod Megiddo. Online Learning with Prior Information. In 20th Conf. on Learning Theory (COLT), pages 499–513, 2007.

[19] O. Maillard. Robust risk-averse stochastic multi-armed bandits. In Proceedings of the 24th International Conference on Algorithmic Learning Theory, pages 218–233. Springer, Berlin, Heidelberg, 2013

[20] J.-Y. Audibert, R. Munos, and Cs. Szepesv´ari. Tuning bandit algorithms in stochastic environments. In Proceedings of the 18th International Conference on Algorithmic Learning Theory, pages 150–165, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[21] J.-Y. Audibert, R. Munos, and Cs. Szepesv´ari. Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. Theoretical Computer Science, 410(19):1876–1902, 2009.

[22] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities: A nonasymptotic theory of independence. OUP Oxford, 2013.

[23] Aleksandrs Slivkins. Contextual bandits with similarity information. J. of Machine Learning Research (JMLR), 15 (1):2533–2568, 2014. Preliminary version in COLT 2011.

[24] Akshay Krishnamurthy, Alekh Agarwal, and Miroslav Dud´ık. Contextual semibandits via supervised learning oracles. In 29th Advances in Neural Information Processing Systems (NIPS), 2016.

[25] Hamsa Bastani, Mohsen Bayati, and Khashayar Khosravi. Mostly exploration-free algorithms for contextual bandits. Management Science, 67(3):1329–1349, 2021. Working paper available on arxiv.org since 2017.

[26] Sampath Kannan, Jamie Morgenstern, Aaron Roth, BoWaggoner, and Zhiwei StevenWu. A smoothed analysis of the greedy algorithm for the linear contextual bandit problem. In Advances in Neural Information Processing Systems (NIPS), 2018.

[27] A. Beygelzimer, J. Langford, L. Li, L. Reyzin, and R.E. Schapire. Contextual bandit algorithms with supervised learning guarantees. In Proceedings of the 15th International Conference on Artificial Intel- ligence and Statistics (AISTATS), JMLR Workshop and Conference Proceedings Volume 15, 2011b.

[28] S. Boucheron, O. Bousquet, and G. Lugosi. Theory of classification: a survey of recent advances. ESAIM: Probability and Statistics, 9: 323–375, 2005.

[29] P. Rusmevichientong and J. Tsitsiklis. Linearly parameterized bandits. Mathematics of Operations Research, 35:395–411, 2010.

[30] V. Perchet and P. Rigollet. The multi-armed bandit problem with covariates. Arxiv preprint arXiv:1110.6084, 2011.

[31] S. Filippi, O. Capp´e, A. Garivier, and C. Szepesv´ari. Parametric bandits: The generalized linear case. In Neural Information Processing Systems (NIPS), 2010.

[32] A. Slivkins. Contextual bandits with similarity information. In Proceedings of the 24th Annual Conference on Learning Theory (COLT), JMLR Workshop and Conference Proceedings Volume 19, 2011.

[33] J. Langford and T. Zhang. The epoch-greedy algorithm for multi-armed bandits with side information. In Advances in Neural Information Processing Systems, pages 817–824. Curran Associates, Inc., 2008.

[34] A. Tewari and S. A. Murphy. From ads to interventions: Contextual bandits in mobile health. In Mobile Health, pages 495–517. Springer, 2017.