Scalable Recommender System for Dota 2

P7

Projekt Rapport

d706e19

Aalborg Universitet - Datalogi

17. februar 2020

Department of Computer Science
Aalborg University
http://www.aau.dk

AALBORG UNIVERSITET
STUDENTERRAPPORT

Title:
P8 - Project

Theme:
Reliable innovative systems

Project Period:
Spring Semester 2020

Project Group:
d706e19

Participant(s):
Andreas Dahl Nielsen
Aryan Mohammadi Landi
Kenneth Kjærgaard Malowanczyk
Lars Svane Jensen
Sebastian Reidar Petersen

Supervisor(s):
Peter Dolog and Manfred Jaeger

Copies: 0

Page Numbers: 14

Date of Completion:
17. februar 2020

Abstract:

This is an abstract

Indhold

Kapitel 1

Introduction

Hey

Bibliografi

[1] E. Charts. (2018). Esports Viewership Stats for 2018, webadr.: https://escharts.com/2018 (bes. 3. dec. 2019).

[2] S. Nordmark og J. Heath. (2019). The top 10 highest prize pools in esports, webadr.: https://dotesports.com/general/news/biggest-prize-pools-esports-14605 (bes. 3. dec. 2019).

[3] Steamcharts. (2019). Steamcharts Dota 2, webadr.: https://steamcharts.com/app/570.

[4] RGP. (2019). Dota 2 Matchmaking Stats, webadr.: https://dota.rgp.io/historic/.

[5] Z. Chen, T. D. Nguyen, Y. Xu, C. Amato, S. Cooper, Y. Sun og M. S. El-Nasr, „The Art of Drafting: A Team-Oriented Hero Recommendation System for Multiplayer Online Battle Arena Games," *CoRR*, årg. abs/1806.10130, 2018. arXiv: 1806.10130. webadr.: http://arxiv.org/abs/1806.10130.

[6] L. Hanke og L. Chaimowicz, „A recommender system for hero line-ups in MOBA games," i *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2017.

[7] Valve. (2019). Dota plus, webadr.: https://www.dota2.com/plus (bes. 3. dec. 2019).

[8] B5810506. (2018). Dota Plus DOES NOT contribute that much to help you. Stop this stupid circlejerk, webadr.: https://www.reddit.com/r/DotA2/comments/83wvb2/dota_plus_does_not_contribute_that_much_to_help/ (bes. 3. dec. 2019).

[9] w3doe. (2018). Is a Dota Plus subscribtion worth it? More so, for a beginner? Webadr.: https://www.reddit.com/r/learndota2/comments/9d5d6x/is_a_dota_plus_subscribtion_worth_it_more_so_for/ (bes. 3. dec. 2019).

[10] D. Gamepedia. (). Dota 2, webadr.: https://dota2.gamepedia.com/Dota_2.

[11] Lemoncream. (). File:Labelled Map 7.20.png, webadr.: https://dota2.gamepedia.com/File:Labelled_Map_7.20.png.

[12] D. Gamepedia. (). Map, webadr.: https://dota2.gamepedia.com/Map.

[13] ——, (13. nov. 2019). Hero Introduction, webadr.: https://dota2.gamepedia.com/Heroes/Introductions.

[14] D. 2. (). All Dota 2 Heroes, webadr.: http://www.dota2.com/heroes/.

[15] D. Gamepedia. (16. mar. 2019). Role, webadr.: https://dota2.gamepedia.com/Role.

[16] dota 2 gamepedia. (27. nov. 2019). Bounty Hunter, webadr.: https://dota2.gamepedia.com/Bounty_Hunter (bes. 12. feb. 2019).

[17] D. Gamepedia. (). Items, webadr.: https://dota2.gamepedia.com/Items.

[18] D. 2. (). All Dota 2 Heroes, webadr.: http://www.dota2.com/items/.

[19] F. Ricci, L. Rokach og B. Shapira, red., *Recommender Systems Handbook*. Springer US, 2015. DOI: 10.1007/978-1-4899-7637-6.

[20] Y. Koren, „Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model," i *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, rk. KDD '08, Las Vegas, Nevada, USA: ACM, 2008, s. 426–434, ISBN: 978-1-60558-193-4. DOI: 10.1145/1401890.1401944. webadr.: http://doi.acm.org/10.1145/1401890.1401944.

[21] G. Takács, I. Pilászy, B. Németh og D. Tikk, „Major Components of the Gravity Recommendation System," *SIGKDD Explor. Newsl.*, årg. 9, nr. 2, s. 80–83, dec. 2007, ISSN: 1931-0145. DOI: 10.1145/1345448.1345466. webadr.: http://doi.acm.org/10.1145/1345448.1345466.

[22] Y. Koren, „Factor in the Neighbors: Scalable and Accurate Collaborative Filtering," *TKDD*, årg. 4, jan. 2010. DOI: 10.1145/1644873.

[23] S. K. Lam og J. Riedl, „Shilling Recommender Systems for Fun and Profit," i *Proceedings of the 13th International Conference on World Wide Web*, rk. WWW '04, New York, NY, USA: ACM, 2004, s. 393–402, ISBN: 1-58113-844-X. DOI: 10.1145/988672.988726. webadr.: http://doi.acm.org/10.1145/988672.988726.

[24] W. Zhou, J. Wen, Y. S. Koh, Q. Xiong, M. Gao, G. Dobbie og S. Alam, „Shilling attacks detection in recommender systems based on target item analysis," *PloS one*, årg. 10, nr. 7, e0130968, 2015.

[25] R. Burke, B. Mobasher, R. Zabicki og R. Bhaumik, „Identifying attack models for secure recommendation," *Beyond Personalization*, årg. 2005, 2005.

[26] R. Burke, B. Mobasher og R. Bhaumik, „Limited knowledge shilling attacks in collaborative filtering systems," *Proceedings of 3rd International Workshop on Intelligent Techniques for Web Personalization (ITWP 2005)*, jan. 2005.

[27] C. A. Williams, B. Mobasher og R. Burke, „Defending recommender systems: detection of profile injection attacks," *Service Oriented Computing and Applications*, årg. 1, nr. 3, s. 157–170, 2007.

[28] M. O'mahony, N. Hurley og G. Silvestre, „An Evaluation of Neighbourhood Formation on the Performance of Collaborative Filtering," *Artificial Intelligence Review*, årg. 21, nr. 3, s. 215–228, jun. 2004, ISSN: 1573-7462. DOI: 10.1023/B:AIRE.0000036256.39422.25. webadr.: https://doi.org/10.1023/B:AIRE.0000036256.39422.25.

[29] B. Mehta og W. Nejdl, „Unsupervised strategies for shilling detection and robust collaborative filtering," *User Modeling and User-Adapted Interaction*, årg. 19, nr. 1, s. 65–97, feb. 2009, ISSN: 1573-1391. DOI: 10.1007/s11257-008-9050-4. webadr.: https://doi.org/10.1007/s11257-008-9050-4.

[30] S. Narkhede. (15. jan. 2019). Scaling Horizontally and Vertically for Databases, webadr.: https://medium.com/@abhinavkorpal/scaling-horizontally-and-vertically-for-databases-a2aef778610c (bes. 14. dec. 2019).

[31] Y. Xin m.fl., „Challenges in recommender systems: scalability, privacy, and structured recommendations," PhD afh., Massachusetts Institute of Technology, 2015.

[32] A. S. Das, M. Datar, A. Garg og S. Rajaram, „Google News Personalization: Scalable Online Collaborative Filtering," i *Proceedings of the 16th International Conference on World Wide Web*, rk. WWW '07, Banff, Alberta, Canada: ACM, 2007, s. 271–280, ISBN: 978-1-59593-654-7. DOI: 10.1145/1242572.1242610. webadr.: http://doi.acm.org/10.1145/1242572.1242610.

[33] R. Gemulla, E. Nijkamp, P. J. Haas og Y. Sismanis, „Large-scale Matrix Factorization with Distributed Stochastic Gradient Descent," i *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, rk. KDD '11, San Diego, California, USA: ACM, 2011, s. 69–77, ISBN: 978-1-4503-0813-7. DOI: 10.1145/2020408.2020426. webadr.: http://doi.acm.org/10.1145/2020408.2020426.

[34] J. Herlocker, J. A. Konstan og J. Riedl, „An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms," *Information Retrieval*, årg. 5, nr. 4, s. 287–310, 2002, ISSN: 1573-7659. DOI: 10.1023/A:1020443909834. webadr.: https://doi.org/10.1023/A:1020443909834.

[35] B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl m.fl., „Item-based collaborative filtering recommendation algorithms.," *Www*, årg. 1, s. 285–295, 2001.

[36] D. M. Allen, „Mean Square Error of Prediction as a Criterion for Selecting Variables," *Technometrics*, årg. 13, nr. 3, s. 469–475, 1971. DOI: 10.1080/00401706.1971.10488811. eprint: https://amstat.tandfonline.com/doi/pdf/10.1080/00401706.1971.10488811. webadr.: https://amstat.tandfonline.com/doi/abs/10.1080/00401706.1971.10488811.

[37] M. Gabbouj og E. J. Coyle, „Minimum mean absolute error stack filtering with structural constraint and goals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, årg. 38, nr. 6, s. 955–968, 1990, ISSN: 0096-3518. DOI: 10.1109/29.56056.

[38] Steam. (2019). Steam Dev Kit, webadr.: https://steamcommunity.com/dev (bes. 2. okt. 2019).

[39] openDota. (2019). openDota Website, webadr.: https://www.opendota.com/ (bes. 2. okt. 2019).

[40] Opendota. (18. dec. 2019). API-keys, webadr.: https://www.opendota.com/api-keys (bes. 18. dec. 2019).

[41] D. L. Poole og A. K. Mackworth, *Artificial Intelligence: Foundations of computational agents (second edition)*. Cambridge University Press, 2017.

[42] V. B. S. Prasath, H. A. A. Alfeilat, A. B. A. Hassanat, O. Lasassmeh, A. S. Tarawneh, M. B. Alhasanat og H. S. E. Salman, „Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier – A Review," 14. aug. 2017. DOI: 10.1089/big.2018.0175. arXiv: 1708.04321v3 [cs.LG].

[43] Netflix. (2009-01-01). The Netflix prize, webadr.: https://www.netflixprize.com (bes. 20. sep. 2019).

[44] A. Töscher og M. Jahrer. (2009-09-05). The BigChaos Solution to the Netflix Grand Prize, webadr.: https://www.netflixprize.com/assets/GrandPrize2009_BPC_BigChaos.pdf (bes. 20. sep. 2019).

[45] S. P. Thandapani. (2019-03-02). Recommendation Systems: Collaborative Filtering using Matrix Factorization — Simplified, webadr.: https://medium.com/sfu-big-data/recommendation-systems-collaborative-filtering-using-matrix-factorization-simplified-2118f4ef2cd3 (bes. 20. sep. 2019).

[46] A. A. Yeung. (2017). Matrix Factorization: A Simple Tutorial and Implementation in Python, webadr.: http://www.albertauyeung.com/post/python-matrix-factorization/ (bes. 3. dec. 2019).

[47] BogoToBogo. (1. jan. 2016). Batch Gradient Descent VS Stochastic Gradient Descent, webadr.: https://www.bogotobogo.com/python/scikit-learn/scikit-learn_batch-gradient-descent-versus-stochastic-gradient-descent.php (bes. 14. dec. 2019).

[48] S. Raschka. (1. jan. 2014). Gradient Descent and Stochastic Gradient Descent, webadr.: http://rasbt.github.io/mlxtend/user_guide/general_concepts/gradient-optimization/?fbclid=IwAR1P0JGjA5h-j_7Nu04NBEqSSUR7Q_J9YCT6XbkKtguaz6PhNj_wfqv8NQI (bes. 17. dec. 2019).

[49] A. Kumar. (2018). Minimax Algorithm in Game Theory, webadr.: https://www.geeksforgeeks.org/minimax-algorithm-in-game-theory-set-1-introduction/ (bes. 28. okt. 2019).

[50] J. Kiefer, „Sequential minimax search for a maximum,“ English, årg. 4, s. 502–, 1953, ISSN: 0002-9939. webadr.: https://sfx.aub.aau.dk/sfxaub?sid=google&auinit=J&aulast=Kiefer&atitle=Sequential+minimax+search+for+a+maximum&title=Proceedings+of+the+American+Mathematical+Society+-+Series+B&volume=4&issue=3&date=1953&spage=502&issn=0002-9939.

[51] O. Wink, W. J. Niessen og M. A. Viergever, „Minimum cost path determination using a simple heuristic function,“ English, årg. 3, s. 998–1001, 2000, ISSN: 1051-4651. DOI: 10.1109/ICPR.2000.903713. webadr.: https://sfx.aub.aau.dk/sfxaub?sid=google&auinit=O&aulast=Wink&atitle=Minimum+cost+path+determination+using+a+simple+heuristic+function&id=doi:10.1109/ICPR.2000.903713&title=2010+20th+International+Conference+on+Pattern+Recognition&volume=3&date=2000&spage=998&issn=1051-4651.

[52] JavaTPoint. (). Java T Point, webadr.: https://www.javatpoint.com/mini-max-algorithm-in-ai (bes. 18. dec. 2019).

[53] P. Borovska og M. Lazarova, „Efficiency of parallel minimax algorithm for game tree search,“ bd. 285, jan. 2007, s. 14. DOI: 10.1145/1330598.1330615.

[54] S. Gelly og D. Silver, „Monte-Carlo tree search and rapid action value estimation in computer Go,“ English, årg. 175, s. 1856–1875, 2011, ISSN: 0004-3702. DOI: 10.1016/j.artint.2011.03.007. webadr.: https://sfx.aub.aau.dk/sfxaub?sid=google&auinit=S&aulast=Gelly&atitle=Monte-Carlo+tree+search+and+rapid+action+value+estimation+in+computer+Go&id=doi:10.1016/j.artint.2011.03.007&title=The+Art+of+Intelligence&volume=175&issue=11&date=2011&spage=1856.

[55] Int8. (2015). Monte Carlo Tree Search – beginners guide, webadr.: https://int8.io/monte-carlo-tree-search-beginners-guide/ (bes. 2. okt. 2019).

[56] S. Narkhede. (26. jun. 2018). Understanding AUC - ROC Curve, webadr.: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5 (bes. 14. dec. 2019).

[57] Scikit-learn. (11. aug. 2019). Scikit-learn, webadr.: https://scikit-learn.org/stable/ (bes. 11. aug. 2019).

[58] J. Heaton. (2017). The number of hidden layers, webadr.: https://www.heatonresearch.com/2017/06/01/hidden-layers.html (bes. 18. nov. 2019).

[59] J. Heaton, *Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks*, rk. Artificial Intelligence for Humans Series. CreateSpace Independent Publishing Platform, 2015, ISBN: 9781505714340. webadr.: https://books.google.dk/books?id=q9mijgEACAAJ.

[60] J. Brownlee. (2019). A Gentle Introduction to the Rectified Linear Unit, webadr.: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/ (bes. 27. nov. 2019).

[61] J. Brownlee. (2019). How to fix vanishing gradient descent Using ReLU, webadr.: https://www.heatonresearch.com/2017/06/01/hidden-layers.html (bes. 27. nov. 2019).

[62] D. P. Kingma og J. Ba, „Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[63] S. Ruder, „An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[64] Scikit-learn. (18. dec. 2019). sklearn.neural-network.MLPClassifier, webadr.: https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html (bes. 18. dec. 2019).

[65] A. Cui, H. Chung og N. Hanson-Holtry, „YASP 3.5 Million Data Dump,"

[66] M. Wang og J. Han, „Collaborative filtering recommendation based on item rating and characteristic information prediction," i *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, 2012, s. 214–217. DOI: 10.1109/CECNet.2012.6201689.

[67] V. Megalooikonomou. (2003). CIS603 Spring 03: Lecture 7, webadr.: https://cis.temple.edu/~vasilis/Courses/CIS603/Lectures/l7.html (bes. 15. dec. 2019).

[68] P. J. Fleming og J. J. Wallace, „How Not to Lie with Statistics: The Correct Way to Summarize Benchmark Results," *Commun. ACM*, årg. 29, nr. 3, s. 218–221, mar. 1986, ISSN: 0001-0782. webadr.: http://doi.acm.org/10.1145/5666.5673.

[69] L. Candillier, F. Meyer og M. Boullé, „Comparing state-of-the-art collaborative filtering systems,“ i *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, Springer, 2007, s. 548–562.

[70] A. Gunawardana og G. Shani, „A survey of accuracy evaluation metrics of recommendation tasks,“ *Journal of Machine Learning Research*, årg. 10, nr. Dec, s. 2935–2962, 2009.

[71] J. Fürnkranz, P. K. Chan, S. Craw, C. Sammut, W. Uther, A. Ratnaparkhi, X. Jin, J. Han, Y. Yang, K. Morik, M. Dorigo, M. Birattari, T. Stützle, P. Brazdil, R. Vilalta, C. Giraud-Carrier, C. Soares, J. Rissanen, R. A. Baxter, I. Bruha, R. A. Baxter, G. I. Webb, L. Torgo, A. Banerjee, H. Shan, S. Ray, P. Tadepalli, Y. Shoham, R. Powers, Y. Shoham, R. Powers, G. I. Webb, S. Ray, S. Scott, H. Blockeel og L. D. Raedt, „Mean Absolute Error,“ i *Encyclopedia of Machine Learning*, Springer US, 2011, s. 652–652. DOI: 10.1007/978-0-387-30164-8_525.

[72] C. Willmott og K Matsuura, „Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,“ *Climate Research*, årg. 30, s. 79–82, 2005. DOI: 10.3354/cr030079.

[73] W. Koehrsen. (3. mar. 2018). Beyond Accuracy: Precision and Recall, webadr.: https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c (bes. 11. jan. 2019).

[74] K. Goldberg, T. Roeder, D. Gupta og C. Perkins, „Eigentaste: A Constant Time Collaborative Filtering Algorithm,“ *Information Retrieval*, årg. 4, nr. 2, s. 133–151, 2001. DOI: 10.1023/a:1011419012209.

[75] T. Cazenave og N. Jouandeau, „On the parallelization of UCT,“ jan. 2007.

[76] G. M. J. B. Chaslot, M. H. M. Winands og H. J. van den Herik, „Parallel Monte-Carlo Tree Search,“ i *Computers and Games*, H. J. van den Herik, X. Xu, Z. Ma og M. H. M. Winands, red., Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, s. 60–71, ISBN: 978-3-540-87608-3.

[77] python. (). Python, webadr.: https://docs.python.org/3/library/multiprocessing.html (bes. 18. dec. 2019).

[78] Scikit-learn. (17. dec. 2019). sklearn.model-selection.GridSearchCVV, webadr.: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html#sklearn.model_selection.GridSearchCV (bes. 17. dec. 2019).

[79] O. Harrison. (10. sep. 2018). Machine Learning Basics with the K-Nearest Neighbors Algorithm, webadr.: https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761 (bes. 18. dec. 2019).

[80] K. Liao. (10. nov. 2018). Prototyping a Recommender System Step by Step Part 1: KNN Item-Based Collaborative Filtering, webadr.: https://towardsdatascience.com/prototyping-a-recommender-system-step-by-step-part-1-knn-item-based-collaborative-filtering-637969614ea (bes. 18. dec. 2019).

[81] Sahanays. (31. jan. 2017). Parallelism in python – Why is using threading module in python a bad idea? Webadr.: https://blogonparallelcomputing.wordpress.com/2017/01/31/parallelism-in-python-why-is-using-the-threading-module-a-bad-idea/ (bes. 18. dec. 2019).

[82] P. Pundir. (26. jul. 2017). Threading vs Multiprocessing in Python, webadr.: https://medium.com/practo-engineering/threading-vs-multiprocessing-in-python-7b57f224eadb (bes. 18. dec. 2019).

[83] Microsoft. (17. dec. 2019). Parallel Programming in .NET, webadr.: https://docs.microsoft.com/en-us/dotnet/standard/parallel-programming/ (bes. 17. dec. 2019).

[84] Oracle. (17. dec. 2019). Parallelism Java Tutorials, webadr.: https://docs.oracle.com/javase/tutorial/collections/streams/parallelism.html (bes. 17. dec. 2019).

[85] U. A. Acar, A. Charguéraud og M. Rainey. (mar. 2016). An Introduction to Parallel Computing in C++, webadr.: https://www.cs.cmu.edu/~15210/pasl.html (bes. 17. dec. 2019).

[86] P. Gribble. (2012). Parallel Programming in C, webadr.: https://gribblelab.org/CBootCamp/A2_Parallel_Programming_in_C.html (bes. 17. dec. 2019).

[87] S. Klabnik og C. Nichols, *The Rust Programming Language*. San Francisco, CA, USA: No Starch Press, 2018, ISBN: 1593278284, 9781593278281.

Bilag A

Acronyms

AI Artificial Intelligence

API Application Programming Interface

OS Operating System

CSV Comma Separated Values

GC Garbage Collection

IDE Integrated Development Environment

JIT Just-in-Time

CBRS Content-Based Recommender Systems

VSM Vector Space Model

CTR Click Through Rate

NDR narrative-driven recommendations

EDR Example-Driven Recommendation

NDCG Normalized Discounted Cumulative Gain

seq2seq Sequence-to-Sequence

MSE Mean Squared Error

HCI Human-Computer Interaction

Dota Defence of the ancients

MCTS Monte Carlo Tree Search

UCT Upper Confidence Tree

UCB Upper Confidence Bound

RMSE Root Mean Squared Error

MAE Mean Absolute Error

MCTS Monte Carlo Tree Search

NMAE Normalized Mean Absolute Error

NN Neural Network

ReLA Rectified Linear Activation Function

ReLU Rectified Linear Unit

Tanh Hyperbolic Tangent

KNN K Nearest Neighbor

AUROC Area Under the Receiver Operating Characteristics

AUC Area Under Curve

Moba Multiplayer Online Battle Arena

SGD Stochastic Gradient Descent

BFGS Broyden-Fletcher-Goldfarb-Shanno

MF Matrix Factorization

LBFGS Limited-memory Broyden-Fletcher-Goldfarb-Shanno

GBDT Gradient Boosted Decision Tree

LR Logistic Regression

ACE Accuracy, Confidence, Effort

GDC Gradient Decent

Adam Adaptive Moment Estimation