December 2025: Top 10 Read Articles in Data Mining & Knowledge Management Process

International Journal of Data Mining & Knowledge Management Process (IJDKP)

ISSN: 2230 - 9608 (Online); 2231 – 006X(print)

https://airccse.org/journal/ijdkp/ijdkp.html

LEARNING CONTEXT FOR TEXT CATEGORIZATION

Y.V. Haribhakta[1] and Dr. Parag Kulkarni[2]

[1]Department of Computer Engineering & I.T. , College of Engineering, Pune,Maharashtra, India

[2]EkLat Labs, Pune, Maharashtra , India

ABSTRACT

This paper describes our work which is based on discovering context for text document categorization.The document categorization approach is derived from a combination of a learning paradigm known as relation extraction and an technique known as context discovery. We demonstrate the effectiveness of our categorization approach using reuters 21578 dataset and synthetic real world data from sports domain. Our experimental results indicate that the learned context greatly improves the categorization performance as compared to traditional categorization approaches.

KEYWORDS

Relation Extraction, Context Discovery, Context Feature Matrix, Context Score

Full Text: https://aircconline.com/ijdkp/V1N6/1611ijdkp02.pdf

REFERENCES

[1] N. H. Yi Guo, Zhiqing Shao, "Automatic text categorization based on Content Analysis withCognitive Situation Model," in Information Sciences. Science Direct, 2010, pp. 613–531.

[2] C.-M. Chen, "Two novel feature selection approaches for web page classification," in ExpertSystems with Applications, Science Direct, 2009, pp. 260–273.

[3] F. S. Giuseppe Attardi, Antonio Gull, "Automatic web page categorization by link and contextanalysis," 2000.

[4] Y.-P. P. C. Jiyuan An, "Keyword extraction for text categorization." IEEE Computer Society,2005, pp. 556–561.

[5] B. F. Andrej Bratko, "Exploiting structural information for SemiStructured documentcategorization," in Information Processing and Management. ACM, 2005

[6] L. W. Y.-F. H. Xiao-Yun Chen, Yi Chen, “Text Categorization based on frequent patterns withterm frequency.” IEEE Computer Society, 2004.

[7] Travis L. Bauer ,David B. Leake , ”Detecting Context-Differentiating Terms Using CompetitiveLearning ”,SIGIR October 2, 2003.

[8] Mohmmed Zaki. Fast Vertical Mining using Diffset. SIGKDD’03, August 2003. Washington, DC,USA Copyright 2003 ACM.

[9] D. B. L. Travis Bauer, “Wordsieve : A method for real-time Context Extraction.” IEEE ComputerSociety.

[10] Fabrizio Sebastiani. Machine Learning in Automated Text Categorization. ACM Computingsurveys ,vol.34, No.1,March 2002, pp.1-47.

[11] V. H. Jihoon Yang, “Feature subset selection using a genetic algorithm,” in ACM ComputingClassification System Categories. ACM, 2000.

[12] J.D. Holt, S.M. Chung. Efficient Mining of Association Rules in Text Databases, CIKM’99,Kansas City, USA,pp.234-242(Nov 1999).

[13] J.S. Park, M.S. Chen and P.S.Yu. Using a Hash-based Method with Transaction trimming forMining Association rules. IEEE Transactions on Knowledge and Data Engineering.Vol9,No.5,Sept/Oct, 1997.

[14] W.W. Cohen and Y. Singer. Context – Sensitive Learning Methods for Text Categorization. Proc19th Int'l ACM SIGIR Conf. Research and Development in Information Retreival , pp.307-315,1996.

[15] D.D. Lewis, R.E. Schapire, J.P. Callan and R. Papka. Training Algorithms for Linear textClassifiers. Proc. 19th Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 298-306, 1996.

[16] R.Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I, Verkamo. Fast discovery ofassociation rules. In Advances in Knowledge Discovery and Data Mining, pages 307–328,1996.

[17] L.S. Larkey and W.B. Croft. Combining Classifiers in Text Categorization. Proc. 19th Int'l ACMSIGIR Conf. Research and Development in Information Retrieval, pp.289-297, 1996.

[18] C. Apte, F. damerau, and S.M. Weiss. Automated Learning of Decision Rules for TextCategorization. ACM Transaction. Information Systems, Vol 12, no. 3, pp 233-251, 1994.

[19] Y. Yang. Expert Network: Effective and Efficient learning from Human Decisions in textcategorization and Retreival. Proc. 17th Int'l ACM SIGIR Conf. Research and Development inInformation retreival, pp.13-22, 1994.

[20] Agrawal R. , Srikanth R. Fast Algorithms for mining association rules VLDB, 1994.

[21] R.Agrawal, T. Imielinski, A. Swami. Mining Associations between Sets of Items inMassive Databases. Proceedings ACM SIGMOD 1993, pp. 207-216.

[22] D.D. Lewis. Feature Selection and Feature Extraction for Text Categorization. Proc. Speech andNatural Language Workshop, pp 212-217, 1992.

[23] R. Uday Kiran and P. Krishna Reddy . An Improved Multiple Minimum Support BasedApproach to Mine Rare Association Rules.

[24] Ayse Ozel and H.Altay. An algorithm for Mining Association Rules using perfect hashing anddatabase pruning. Bilkent University, Department of Computer Engineering, Ankara, Turkey.

[25] www.daviddlewis.com/resources/testcollections/reuters21578.

FOCUSED WEB CRAWLING USING DECAY CONCEPT AND GENETIC PROGRAMMING

Mahdi Bazarganigilani1, Ali Syed2 and Sandid Burki3

Faculty of Business, Charles Sturt University, Melbourne, Australia

ABSTRACT

The ongoing rapid growth of web information is a theme of research in many papers. In this paper, we introduce a new optimized method for web crawling. Using genetic programming enhances the accuracy of simialrity measurement. This measurement applies to different parts of the web pages including thetitle and the body. Consequently, the crawler uses such optimized similarity measurement to traverse the pages .To enhance the accuracy of crawling, we use the decay concept to limit the crawler to theeffective web pages in accordance to search criteria. The decay measurements give every page a score according to the search criteria. It decreases while traversing in more depth. This value could be revised according to the similarity of the page to the search criteria. In such case, we use three kinds of measurement to set the thresholds. The results show using Genetic programming along the dynamic decay thresholds leads to the best accuracy.

KEYWORDS

Focused Web Crawler; Genetic Programming; Decay Concept; Similarity Space Model

Full Text: https://aircconline.com/ijdkp/V1N1/1111ijdkp01.pdf

REFERENCES

[1] A. Gulli, A. Signorini, “The Indexable web is more than 11.5 billion pages”, In Proceedings of the 14th international conference on World Wide Web, pp. 902- 903, ACM Press, 2005.

[2] Internet Metrics and Statistics Guide: Size and Shape, http://caslon.com.au/metricsguide13.htm, Version of 2003 "Evaluating Crawling Efficiency Using Different Weighting Schemes with Regional Crawler", P. Chubak,

[3] M. Shokouhi, Proceedings of IEEE 4th International Conference o Intelligent Systems Design and Applications (ISDA2004), Budapest, Hugary, 2004.

[4] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval, Addison-Wesley, 1999.

[5] M. Khalilian, K. SheikhEsmaili, M. Neshati, H. Abolhassani, "Boundy Threshold Controling Using Decay Concept In Focused Crawling ", Thirteenth National CSI Computer Conference, Kish Islan, Iran, March 2008.

[6] Bra, P.D., Houben, G., Kornatzky, Y., and Post, R., Information Retrieval in Distributed Hypertexts. in Proceedings of the 4th RIAO Conference. p. 481-491. 1994. New York.

[7] M. Ehrig, A. Maedche. Ontology-focused Crawling of Web Documents, In Proceedings of the 2003 ACM symposium on Applied computing.

[8] Yuxin Chen. A novel hybrid focused crawling algorithm to build domain-specific collections. PhD thesis,

United States – Virginia, 2007.

[9] Qin, J. and Chen, H., Using Genetic Algorithm in Building Domain-Specific Collections: An Experiment in the Nanotechnology Domain. in Proceedings of the 38th Annual Hawaii International Conference on System Sciences - HICSS 05. p. 102.2. 2005. Hawaii, USA.

[10] F. Menczer and G. Pant and P. Srinivasan. Topic-driven crawlers: Machine learning issues, ACMTOIT, Submitted, 2002.

[11] S. Chakrabarti,M. van den Berg and B. Dom. Focused Crawling: A New Approach to Topic- Specific Web Resource Discovery, In Proceedings of the 8th International WWW Conference, Toronto, Canada, May 1999.

[12] D.Maghesh Kumar,, (2010) "Automatic Induction of Rule Based Text Categorization", International Journal of Computer Science & Information Technology (IJCSIT), Vol 2, No 6, December 2010.

[13] J. Cho, H. Garcia-Molina, L. Efficient Crawling Through URL Ordering, Page. In Proceedings of the 7th International WWW Conference, Brisbane, Australia, April 1998.

[14] L. Page, S. Brin, R. Motwani, T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web, Stanford Digital Library Technologies Project.

[15] D. Bergmark and C. Lagoze and A. Sbityakov. Focused Crawls, Tunneling, and Digital Libraries.

[16] M. Jamali, H. Sayyadi, B. Bagheri Hariri and H. Abolhassani. A Method for Focused Crawling Using Combination of Link Structure and Content Similarity, WI/IEEE/ACM 2006. Hong Kong, 2006.

[17] M. Diligenti, F. Coetzee, S. Lawrence, C. Giles and M. Gori. Focused Crawling Using Context Graphs, In Proceedings of the 26th International Conference on Very Large Databases (VLDB 2000), Cairo, Egypt, September2000.

[18] Castillo, M.D.D. and Serrano, J.I., A multistrategy approach for digital text categorization from imbalanced documents. SIGKDD, 2004. 6(1): p. 70-79.

[19] Zhang, B., Chen, Y., Fan, W., Fox, E.A., Gonçalves, M.A., Cristo, M., and Calado, P., Intelligent Fusion of Structural and Citation-Based Evidence for Text Classification. in Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 667-668. 2005. Salvador, Brazil.

[20] Zhang, B., Chen, Y., Fan, W., Fox, E.A., Gonçalves, M.A., Cristo, M., and Calado, P., Intelligent GP Fusion from Multiple Sources for Text Classification. in Proceedings of the 14th Conference on Information and Knowledge Management. p. 477-484. 2005. Bremen, Germany.

[21] Zhang, B., Gonçalves, M.A., Fan, W., Chen, Y., Fox, E.A., Calado, P., and Cristo, M., Combining Structure and Citation-Based Evidence for Text Classification. in Proceedings of the 13th Conference on Information and Knowledge Management. p. 162-163. 2004. Washington D.C., USA.

[22] Zhang, B., Gonçalves, M.A., Fan, W., Chen, Y., Fox, E.A., Calado, P., and Cristo, M., A Genetic Programming Approach for Combining Structural and Citation-Based Evidence for Text Classification in Web Digital Libraries, in Soft Computing in Web Information Retrieval: Models and Applications. 2006: p. 65-83.

[23] Salton, G., Automatic Text Processing. 1989, Boston, Massachusetts, USA: Addison-Wesley.

[24] Kleinberg, J.M., Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM), 1999. 46(5): p. 604-632.

[25] Bergmark, D., Collection Synthesis. in Proceedings of the 2nd ACM/IEEE-CS joint conference on digital libraries p. 253-262. 2002. Portland, Oregon, USA.

[26] Dean, J. and Henzinger, M.R., Finding Related Pages in the World Wide Web. in Proceedings of the 8th International WWW Conference. p. 1467-1479. 1999. Toronto, Canada.

[27] Kitsuregawa, M., Toyoda, M., and Pramudiono, I., WEB Community Mining and WEB Log Mining: Commodity Cluster Based Execution. in Proceedings of the 13th Australasian Database Conference. p. 3-10. 2002. Melbourne, Australia.

[28] Salton, G. and Buckley, C., Term-weighting approaches in automatic text retrieval. IPM, 1988. 24(5): p. 513-523.

[29] Yang, Y., Expert network: effective and efficient learning from human decisions in text categorization and retrieval. in Proceedings of 17th ACM International Conference on Research and Development in Information Retrieval - SIGIR 94. p. 13-22. 1994. Dublin, Ireland.

[30] Koza, J.R., Genetic programming: On the programming of computers by natural selection. 1992, Cambridge, MA, USA: MIT Press.

[31] Fan, W., Fox, E.A., Pathak, P., and Wu, H., The effects of fitness functions on genetic programmingbased ranking discovery for web search. JASIST, 2004. 55(7): p. 628-636.

[32] DMOZ, Directory Mozilla, http://www.dmoz.org,.

[33] Robertson, S.E., Walker, S., and Beaulieu, M.M., Okapi at TREC-4. in TREC-4. p. 73-96. 1995.

[34] Joachims, T., Cristianini, N., and Shawe-Taylor, J., Composite kernels for hypertext categorisation. in Proceedings of 18th International Conference on Machine Learning - ICML 01. p. 250-257. 2001. Williams College, USA.

[35] Cleverdon, C. W. and Mills, J. The testing of index language devices. Aslib Proceeding. 15, 4, 106– 130, 1963

A WEB REPOSITORY SYSTEM FOR DATA MINING IN DRUG DISCOVERY

Jiali Tang, Jack Wang and Ahmad Reza Hadaegh

Department of Computer Science and Information System, California State University San Marcos, San Marcos, USA

ABSTRACT

This project is to produce a repository database system of drugs, drug features (properties), and drug targets where data can be mined and analyzed. Drug targets are different proteins that drugs try to bind to stop the activities of the protein. Users can utilize the database to mine useful data to predict the specific chemical properties that will have the relative efficacy of a specific target and the coefficient for each chemical property. This database system can be equipped with different data mining approaches/algorithms such as linear, non-linear, and classification types of data modelling. The data models have enhanced with the Genetic Evolution (GE) algorithms. This paper discusses implementation with the linear data models such as Multiple Linear Regression (MLR), Partial Least Square Regression (PLSR), and Support Vector Machine (SVM).

KEYWORDS

Data Mining, Drug Discovery, Drug Description, Chemoinformatics, and Web Application

Full Text: **https://aircconline.com/ijdkp/V10N1/10120ijdkp01.pdf**

REFERENCES

[1] Ko, Gene, Reddy, Srinivas, Garg, Rajni, Kumar, Sunil, & Hadaegh, Ahmad, (2012) "Computational Modelling Methods for QSAR Studies on HIV-1 Integrase Inhibitors (2005-2010),". Curr Comput Aided Drug Des. Vol. 8, No 4, pp 255-270.

[2] Thakor, Falguni, Hadaegh, Ahmad, & Zhang, Xiaoyu, (2017), " Comparative study of Differential Evolutionary-Binary Particle Swarm Optimization (DE-BPSO) algorithm as a feature selection technique with different linear regression models for analysis of HIV-1 Integrase Inhibition features of Aryl β-Diketo Acids", Proceedings of 9th International Conference on Bioinformatics and Computational Biology, Honolulu, Hawaii, USA, ISBN: 978–1–943436–07–1, pp 179-184.

[3] Kane Ian, & Hadaegh Ahmad, "Non-linear Quantitative Structure-Activity Relationship (QSAR) Models for the Prediction of HIV Drug Performance", (2015), 24th International Conference on Software Engineering and Data Engineering, pp 63-68. Vol 1, ISBN: 9781510812277, San Diego, CA.

[4] Galvan Richard, Kashani, Maninatalsadat, & Hadaegh, Ahmad, "Improving Pharmacological Research of HIV-1 Integrase Inhibition Using Differential Evolution-Binary Particle Swarm Optimization and Non-Linear Adaptive Boosting Random Forest Regression",(2015), IEEE International Workshop on Data Integration and Mining San Francisco, Information Reuse and Integration (IRI), IEEE International Conference, pp 485-490, DOI: 10.1109/IRI.2015.80. INSPEC Accession Number: 15556631. San Francisco, CA.

[5] Kashani, Maninatalsadat, Galvan Richard, & Hadaegh Ahmad, "Improving the Feature Selection for the

Development of Linear Model for Discovery of HIV-1 Integrase Inhibitors", (2015) ABDA'15 International Conference on Advances in Big Data Analytics. In Proceeding of the 2015 International Conferences on Advances on Big Data Analyses, pp 150-154. ISBN: 1-60132-411-1, Las Vegas, Nevada.

[6] Ko, Gene, Garg, Rajni, Kumar, Sunil, Kumar, Bailey, Barbara, & Hadaegh Ahmad, "A Hybridized Evolutionary Algorithm for Feature Selection of Chemical Descriptors for Computational QSAR Modeling of HIV-1 Integrase Inhibitors", (2013), Computational Science Curriculum Development Forum and Applied Computational Science and Engineering Student Support for Industry, San Diego State University.

[7] Ko, Gene, Garg, Rajni, Kumar, Sunil, Bailey, Barbara, & Hadaegh Ahmad, "Differential EvolutionBinary Particle Swarm Optimization for the Analysis of Aryl b-Kiketo Acids for HIV-1 Integres Inhibition, (2012), WCCI 2012 IEEE World Congress on Computational Intelligence. Brisbane Australia, pp 1849-1855.

[8] Ko, Gene, Reddy, Srinivas, Kumar, Kumar, Bailey, Barbara, Garg, Rajni, & Hadaegh, Ahmad, "Evolutionary Computational Modelling of β-Diketo Acids for Virtual Screening of HIV-1 Integrase Inhibitors", (2012), IEEE World Congress on Computational Intelligence, Brisbane, Australia.

[9] Ko, Gene, Reddy, Srinivas, Kumar, Kumar, Garg, Rajni, & Hadaegh, Ahmad "Evolutionary Computational Modelling of β-Diketo Acids for Virtual Screening of HIV-1 Integrase Inhibitors", (2012), 243rd National Meeting of the American Chemical Society, San Diego, CA.

[10] Gonzales, Miguel, Turner, Chris, Ko, Gene, & Hadaegh, Ahmad, "Binary Particle Swarm Optimization Model of Dimeric Aryl Diketo Acid Inhibitors for HIV-1 Integrase" (2012), 243rd National Meeting of the American Chemical Society, San Diego, CA.

[11] Ko, Gene, Reddy, Srinivas, Kumar, Sunil, Garg, Rajni, & Hadaegh, Ahmad, "Analysis of HIV-1 Integrase Inhibitors Using Computational QSAR Modelling", (2012), Computational Science Curriculum Development Forum and Applied Computational Science and Engineering Student Support for Industry, San Diego State University.

[12] Garg Rajni, Reddy Srinivas, Zhang Xiaoyu, & Hadaegh Ahmad, "MUT-HIV: Mutation database of HIV proteases", (2007), American Chemical Society (ACS) 234th National Meeting & Exposition, Boston, MA USA CINF 42.

[13] MLR: http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm

[14] PLSR: https://www.mathworks.com/help/stats/plsregress.html

[15] https://techdifferences.com/difference-between-descriptive-and-predictive-data-mining.html

[16] Zhong et al. Artificial intelligence in drug design. Sci China Life Sci. 2018 Jul 18. doi: 10.1007/s11427-018-9342-2. [Epub ahead of print]

[17] Varsou Dimitra-Danai, Nikolakopoulos, Spyridon, Tsoumanis Andreas, Melagraki Georgia, & Afantitis, Antreas, "New Cheminformatics Platform for Drug Discovery and Computational Toxicology", (2018), Methods Mol Biol. 2018; 1800:287-311. doi: 10.1007/978-1-4939-7899-1_14

[18] Ekins, Sean, Clark, Alex, Dole, Krishna, Gregory, Kellan, Mcnutt, Andrew, Spektor, Anna, Weatherall, Charlie, & Litterman, Nadia, "Data Mining and Computational Modeling of HighThroughput Screening Datasets", (2018), Methods Mol Biol, 1755:197-221. doi: 10.1007/978-1- 4939-7724-6_14.

[19] Sam Elizabeth, & Athri Prashanth, "Web-based drug repurposing tools: a survey. Brief Bioinform", (2017), Oct 6. doi: 10.1093/bib/bbx125. [Epub ahead of print].

[20] Kaur, Charanpreet, & Bhardwaj, Shweta, "DRUG Discovery Using Data Mining International Journal of Information and Computation Technology", (2014), ISSN 0974-2239 Volume 4, Number 4, pp 335-342 © International Research Publications House http://www. irphouse.com /ijict.htm

[21] Minaei-Bidgoli, Behrouz, & Punch, William, "Using Genetic Algorithms for Data Mining Optimization in an Educational Web-Based System, (2003), Genetic Algorithms Research and Applications Group (GARAGe) Department of Computer Science & Engineering Michigan State University 2340 Engineering Building East Lansing, MI 48824.

[22] https://chm.kode-solutions.net/products_dragon.php

[23] AWS LightSail: https://aws.amazon.com/lightsail/?nc2=h_ql_prod_fs_ls [24] AWS EC2 Server: https://aws.amazon.com/ec2

APPLICATION OF SPATIOTEMPORAL ASSOCIATION RULES ON SOLAR DATA TO SUPPORT SPACE WEATHER FORECASTING

Carlos Roberto Silveira Junior1 , José Roberto Cecatto2 , Marilde Terezinha Prado Santos1 and Marcela Xavier Ribeiro1

1 Department of Computing, Federal University of São Carlos, São Carlos, Brazil 2 National Institute of Space Research, São José dos Campos, Brazil

ABSTRACT

It is well known that solar energetic phenomena influence the Space Weather, in special those directed to the Earth environment. In this context, the analysis of Solar Data is a challenging task, particularly when are composed of Satellite Image Time Series (SITS). It is a multidisciplinary domain that generates a massive amount of data (several Gigabytes per year). It includes image processing, spatiotemporal characteristics, and the processing of semantic data. Aiming to enhance the SITS analysis, we propose an algorithm called "Miner of Thematic Spatiotemporal Associations for Images" (MiTSAI), which is an extractor of Thematic Spatiotemporal Association Rules (TSARs) from Solar SITS. Here, a description is given about the details of the modern algorithm MiTSAI, which is an extractor of Thematic Spatiotemporal Association Rules (TSARs) from solar Satellite Image Time Series (SITS). In addition, its adaptation to the Space Weather and discussion about the specific use in favor of forecasting activities are presented. Finally, some results of its application specifically to solar flare forecasting are also presented. MiTSAI has to extract interesting new patterns compared with the art-state algorithms.

KEYWORDS

Satellite Image Time Series; Thematic Spatiotemporal Association Rules; Space Weather Patterns.

Full Text: **https://aircconline.com/ijdkp/V10N2/10220ijdkp01.pdf**

REFERENCES

[1] T Abirami-Kongu, P Thangaraj, and P Priakanth-Kongu. Wireless sensor networks fault identification using data association. Journal of Computer Science, 8(9):1501–1505, 2012.

[2] Sultan Alamri, David Taniar, and Maytham Safar. A taxonomy for moving object queries in spatialdatabases. Future Generation Computer Systems, 37:232 – 242, 2014. ISSN 0167-739X. doi: https://doi.org/10.1016/j.future.2014.02.007.

[3] Shadi A. Aljawarneh, Radhakrishna Vangipuram, Veereswara Kumar Puligadda, and Janaki Vinjamuri. G-spamine: An approach to discover temporal association patterns and trends in internet of things. Future Generation Computer Systems, 2017. ISSN 0167-739X. doi: https://doi.org/10.1016/ j.future.2017.01.013.

[4] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. pages 404–417, 2006.

[5] Monica G Bobra and Sebastien Couvidat. Solar flare prediction using sdo/hmi vector magnetic field data with a machine-learning algorithm. The Astrophysical Journal, 798(2):135–172, 2015.

[6] I. Burbey and T.L. Martin. A survey on predicting personal mobility. International Journal of Pervasive Computing and Communications, 8(1):5 – 22, 2012.

[7] M. Chen, S. Mao, and Y. Liu. Big data: A survey. Mobile Networks and Applications, 19(2):171–209,2014.

[8] Paolo Compieta, Sergio Di Martino, Michela Bertolotto, Filomena Ferrucci, and T Kechadi. Exploratory spatio-temporal data mining and visualization. Journal of Visual Languages & Computing, 18(3):255–279, 2007.

[9] G Fang and Y Wu. Frequent spatiotemporal association patterns mining based on granular computing. Informatica (Slovenia), 37(4):443–453, 2013.

[10] A. Hana, Y.T. Sami, and F. Sami. Mining spatiotemporal associations using queries. 2012 International Conference on Information Technology and e-Services, ICITeS 2012, 2012

[11] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6):610–621, Nov 1973. ISSN 0018-9472. doi: 10.1109/TSMC.1973.4309314.

[12] J. Huo, J. Zhang, and X. Meng. On co-occurrence pattern discovery from spatio-temporal event stream. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8181 LNCS(PART 2):385–395, 2013.

[13] J. Kawale, S. Liess, V. Kumar, U. Lall, and A. Ganguly. Mining time-lagged relationships in spatiotemporal climate data. pages 130–135, 2012.

[14] Xiangjie Kong, Zhenzhen Xu, Guojiang Shen, Jinzhong Wang, Qiuyuan Yang, and Benshi Zhang. Urban traffic congestion estimation and prediction based on floating car trajectory data. Future Generation Computer Systems, 61:97 – 107, 2016. ISSN 0167-739X. doi: https://doi.org/10.1016/j.future. 2015.11.013.

[15] T.C.W. Landgrebe, A. Merdith, A. Dutkiewicz, and R.D. Mafaler. Relationships between palaeogeography and opal occurrence in australia: A data-mining approach. Computers and Geosciences, 56: 76–82, 2013.

[16] Angela Lausch, Andreas Schmidt, and Lutz Tischendorf. Data mining and linked open data: New perspectives for data analysis in environmental research. Ecological Modelling, 295(0):5 – 17, 2015. ISSN 0304-3800. doi: http://dx.doi.org/10.1016/j.ecolmodel.2014.09.018.

[17] A. Madraky, Z.A. Othman, and A.R. Hamdan. Analytic methods for spatio-temporal data in a natureinspired data model. International Review on Computers and Software, 9(3):547–556, 2014.

[18] A.a Mohan and P.Z.b Revesz. Applications of spatio-temporal data mining to north platter river reservoirs. ACM International Conference Proceeding Series, pages 306–309, 2014.

[19] NOAA. www.solarmonitor.org, April 2016. Last Access April 13, 2016.

[20] K.G. Pillai, R.A. Angryk, J.M. Banda, M.A. Schuh, and T. Wylie. Spatio-temporal co-occurrence pattern mining in data sets with evolving regions. Proceedings - 12th IEEE International Conference on Data Mining Workshops, ICDMW 2012, pages 805–812, 2012.

[21] K.G.a Pillai, R.A.b Angryk, and B.b Aydin. A filter-and-refine approach to mine spatiotemporal cooccurrences. pages 104–113, 2013.

[22] Vangipuram Radhakrishna, Shadi A. Aljawarneh, P.V. Kumar, and V. Janaki. A novel fuzzy similarity measure and prevalence estimation approach for similarity profiled temporal association pattern mining. Future Generation Computer Systems, 2017. ISSN 0167-739X. doi: https://doi.org/10.1016/j.

future.2017.03.016.

[23] K Venkateswara Rao, A Govardhan, and KV Chalapati Rao. Spatiotemporal data mining: Issues, tasks and applications. International Journal of Computer Science & Engineering Survey (IJCSES) Vol, 3:39–52, 2012.

[24] Md. Mamunur Rashid, Iqbal Gondal, and Joarder Kamruzzaman. A technique for parallel sharefrequent sensor pattern mining from wireless sensor networks. Procedia Computer Science, 29(0):124 – 133, 2014. ISSN 1877-0509. doi: http://dx.doi.org/10.1016/j.procs.2014.05.012. 2014 International Conference on Computational Science.

[25] Md.M. Rashid, I. Gondal, and J. Kamruzzaman. Mining associated sensor patterns for data stream of wireless sensor networks. pages 91–98, 2013.

[26] Marcela Xavier Ribeiro, Agma J. M. Traina, and Caetano Traina, Jr. A new algorithm for data discretization and feature selection. In Proceedings of the 2008 ACM symposium on Applied computing, SAC '08, pages 953–954, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-753-7. doi: 10.1145/1363686.1363905.

[27] James A. Rodger. Toward reducing failure risk in an integrated vehicle health maintenance system: A fuzzy multi-sensor data fusion kalman filter approach for ivhms. Expert Systems with Applications, 39(10):9821 – 9836, 2012. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2012.02.171.

[28] W. Sammouri, E. Cafame, L. Oukhellou, and P. Aknin. Mining floating train data sequences for temporal association rules within a predictive maintenance framework. Lecture Notes in Computer subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7987 LNAI:112–126, 2013.

[29] T. Scheffer. Finding association rules that trade support optimally against confidence. pages 9: 381– 395, 1995.

[30] Carlos Roberto Silveira-Junior. github.com/carlossilveirajr/mitsai, April 2017. Last Access April 17, 2017.

[31] Carlos Roberto Silveira-Junior, Danilo Codeco Carvalho, Marilde Terezinha Prado Santos, and Marcela Xavier Ribeiro. Incremental mining of frequent sequences in environmental sensor data. In The Twenty-Sixth International FLAIRS Conference (2015), pages 452–455, .

[32] Carlos Roberto Silveira-Junior, Marilde Prado Santos, and Marcela Ribeiro. Stretchy time pattern mining: A deeper analysis of environment sensor data. pages 468–473, .

[33] Carlos Roberto Silveira-Junior, Marcela Xavier Ribeiro, and Marilde Terezinha Prado Santos. A flexible architecture to integrate the solar satellite image time series data - the setl architecture. Pages 1–14, 2017. No publish by the time of this paper preparation.

[34] Olga Spatenkovaa and Kirsi Virrantausb. Discovering spatio-temporal relationships in the distribution of building fires. Fire Safety Journal, 62, Part A:49 – 63, 2013. ISSN 0379-7112. doi: http://dx. doi.org/10.1016/j.firesaf.2013.07.001. Special Issue on Spatial Analytical Approaches in Urban Fire Management.

[35] Fenzhen Su, Chenghu Zhou, and Wenzhoung Shi. Geoevent association rule discovery model based on rough set with marine fishery application. In Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International, volume 2, pages 1455–1458 vol.2, Sept 2004. doi: 10.1109/IGARSS.2004.1368694.

[36] Edi Winarko and John F. Roddick. Armada: An algorithm for discovering richer relative temporal association rules from interval-based data. Data & Knowledge Engineering, 63(1):76 – 90, 2007 ISSN 0169-023X. doi: http://dx.doi.org/10.1016/j.datak.2006.10.009. Data Warehouse an Knowledge Discovery, 7th International Congress on Data Warehouse and Knowledge Discovery.

[37] J.S. Yoo and M. Bow. Mining spatial colocation patterns: A different framework. Data Mining and Knowledge Discovery, 24(1):159–194, 2012.

[38] B. Zaragoza, A. Rabasa, J.J. Rodriguez-Sala, J.T. Navarro, A. Belda, and A. Ramon. Modelling farmland abandonment: A study combining gis and data mining techniques. volume 155, pages 124 – 132, 2012. doi: http://dx.doi.org/10.1016/j.agee.2012.03.019.

PETROCHEMICAL PRODUCTION BIG DATA AND ITS FOUR TYPICAL APPLICATION PARADIGMS

Hu Shaolin* , Zhang Qinghua, Su Naiquan, and Li Xiwu
Guangdong University of Petrochemical Technology, Maoming, Guangdong, China

ABSTRACT

In recent years, the big data has attracted more and more attention. It can bring us more information and broader perspective to analyse and deal with problems than the conventional situation. However, so far, there is no widely acceptable and measurable definition for the term “big data”. For example, what significant features a data set needs to have can be called big data, and how large a data set is can becalled big data, and so on. Although the "5V" description widely used in textbooks has been tried to solve the above problems in many big data literatures, "5V" still has significant shortcomings and limitations, and is not suitable for completely describing big data problems in practical fields such as industrial production. Therefore, this paper creatively puts forward the new concept of data cloud and the data cloud-based "3M" descriptive definition of big data, which refers to a wide range of data sources (Multisource), ultra-high dimensions (Multi-dimensional) and a long enough time span (Multi-spatiotemporal). Based on the 3M description of big data, this paper sets up four typical application paradigms for the production big data, analyses the typical application of four paradigms of big data, and lays the foundation for applications of big data from petrochemical industry.

KEYWORDS

Big Data, Paradigms, Industrial Big Data.

Full Text: https://aircconline.com/ijdkp/V11N4/11421ijdkp02.pdf

REFERENCES

[1] Li Ning(2019).Artificial Intelligence Paradigm in Big Data Era. China Computer & Communication, 8:104-105.

[2] Jarosław W, Jarosław J, Paweł Z(2019), Generalised framework for Multi-criteria Method Selection. Omega,86:107–124.

[3] Manyika J, Chui M, Brown B, Bughin J.,et al(2017). Big data: The Next Frontier for Innovation, Competition, and productivity. McKinsey Global Institute

[4] Abbass H. A, Leu G,Merrick K(2016).A Review of Theoretical and Practical Challenges of Trusted Autonomy in Big Data. Theoretical Foundations for Big Data Applications: Challenges and Opportunities.

[5] Ammu N, Irfanuddin M. Big data challenges. International Journal of Advanced Trends in Computer Science and Engineering, 2013,2(1), 613-615

[6] Rabhi L, Falin N, Afraites A,et al(2019).Procedia Computer Science,16th International Conf. on Mobile Systems and Pervasive Computing, v155, pp:599-605.

[7] Mukherjee S, Shaw R(2016). Big Data-Concepts, Applications, Challenges and Future Scope. Wikipedia, https://en.wikipedia.org/wiki/Google_Cloud_Platform

[8] Tanya Garg;Surbhi Khullar(2020). Big Data Analytics: Applications, Challenges & Future Directions. 8th International Confer on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). pp:923-928,4-5 June 2020, Noida, India.

[9] Nojood Aljehane(2020).Big Data Analytics: Challenges and Opportunities. International Confer on Computing and Information Technology (ICCIT-1441), pp:1-4, 9-10 Sept. 2020, Tabuk, Saudi Arabia.

[10] Ma Z, Yang L, Zhang, Q(2021). Support Multimode Tensor Machine for Multiple Classification on Industrial Big Data. IEEE Transactions on Industrial Informatics,17(5):3382-3390.

[11] Sakineti S; Prabhu C (2018). Protagonist of Big Data and Predictive Analytics using Data Analytics. International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), pp:276-279,21-22 Dec. 2018, Belgaum,India.

[12] Zhou X, Hu Y, Liang W,et al(2021). Variational LSTM Enhanced Anomaly Detection for Industrial Big Data. IEEE Transactions on Industrial Informatics,17(5): 3469-3477.

[13] Sfaxi L,Ben A,Mohamed M(2020), DECIDE: An Agile Event-and-Data Driven Design Methodology for Decisional Big Data Projects. Knowledge Engineering, volume 130, DOI: 10.1016/j.datak.2020.101862.

[14] Battistelli G, Tesi, P(2021). Classification for Dynamical Systems: Model-Based and Data-Driven Approaches.IEEE Transactions on Automatic Control, 66(4):1741-1748.

SEMANTICS GRAPH MINING FOR TOPIC DISCOVERY AND WORD ASSOCIATIONS

Alex Romanova

Melenar, LLC, McLean, VA, USA

ABSTRACT

Big Data creates many challenges for data mining experts, in particular in getting meanings of text data. It is beneficial for text mining to build a bridge between word embedding process and graph capacity to connect the dots and represent complex correlations between entities. In this study we examine processes of building a semantic graph model to determine word associations and discover document topics. We introduce a novel Word2Vec2Graph model that is built on top of Word2Vec word embedding model. We demonstrate how this model can be used to analyze long documents, get unexpected word associations and uncover document topics. To validate topic discovery method we transfer words to vectors and vectors to images and use CNN deep learning image classification.

KEYWORDS

Graph Mining, Semantics, Topics Discovery, Word Associations, Deep Learning, Transfer Learning, CNN Image Classification.

Full Text: **https://aircconline.com/ijdkp/V11N4/11421ijdkp01.pdf**

REFERENCES

[1] Alex Thomas (2020) Natural Language Processing with Spark NLP, O'Reilly Media, Inc.

[2] T Mikolov & I Sutskever & K Chen & GS Corrado & J Dean, (2013) "Distributed representations of words and phrases and their compositionality", Neural information processing systems.

[3] Andrew Cattle and Xiaojuan Ma, (2017) "Predicting Word Association Strengths", 2017Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1283–1288.

[4] Bill Chambers &Matei Zaharia (2018) Spark: The Definitive Guide: Big Data Processing Made Simple, O'Reilly Media, Inc.

[5] Jurij Leskovec & Marko Grobelnik & Natasa Milic-Frayling, (2004). "Learning Substructures of Document Semantic Graphs for Document Summarization", LinkKDD 2004

[6] Juan Martinez-Romo & Lourdes Araujo & Andres Duque Fernandez, (2016). "SemGraph: Extracting Keyphrases Following a Novel Semantic Graph-Based Approach", Journal of the Association for Information Science and Technology, 67(1):71–82, 2016

[7] Long Chen and Joemon M Jose and Haitao Yu and Fajie Yuan, (2017) "A Semantic Graph-Based Approach for Mining Common Topics from Multiple Asynchronous Text Streams", 2017 International World Wide Web Conference Committee (IW3C2)

[8] Michael Thelwall, (2021) "Word Association Thematic Analysis: A Social Media Text volume 13, pages i-111

[9] Andrew Cattle and Xiaojuan Ma, (2017) “Predicting Word Association Strengths”, 2017 Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1283–1288

[10] Matan Zuckerman & Mark Last, (2019) “Using Graphs for Word Embedding with Enhanced Semantic Relations”, Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13).

[11] Long Chen & Joemon M Jose & Haitao Yu & Fajie Yuan & Dell Zhang, (2016). "A Semantic Graph based Topic Model for Question Retrieval in Community Question Answering", WSDM’16

[12] Jintao Tang & Ting Wang & Qin Lu Ji & Wang & Wenjie Li, (2011)."A Wikipedia Based Semantic Graph Model for Topic Tracking in Blogosphere", IJCAI’11

[13] Stavros Souravlas & Angelo Sifaleras & M Tsintogianni & Stefanos Katsavounis, (2021). "A classification of community detection methods in social networks: A survey", International Journal ofGeneral Systems 50(1):63-91

[14] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, PierreAlain Muller: Deep learning for time series classification: a review. Data Min Knowl Disc 33, 917–963 (2019)

[15] Nima Hatami, Yann Gavet, Johan Debayle: Classification of time-series images using deep convolutional neural networks Conference: Tenth International Conference on Machine Vision (ICMV 2017).

[16] Zhiguang Wang, Tim Oates: Encoding Time Series as Images for Visual Inspection andClassification Using Tiled Convolutional Neural Networks. Association for the Advancement of Artificial Intelligence (www.aaai.org)(2015)

[17] Zhiguang Wang, Weizhong Yan, Tim Oates: Time series classification from scratch with deep neural networks: A strong baseline. International Joint Conference on Neural Networks (IJCNN)(2017)

[18] "Sparkling Data Ocean - Data Art and Science in Spark”, http://sparklingdataocean.com/

[19] Yoav Goldberg&Graeme Hirst (2017) Neural Network Methods in Natural Language Processing, Morgan & Claypool Publishers.

[20] "Word2Vec Model Training", http://sparklingdataocean.com/2017/09/06/w2vTrain/

[21] "Introduction to Word2Vec2GraphModel”,http://sparklingdataocean.com/2017/09/17word2vec2graph

[22] Alex Romanova, (2020) “Building Knowledge Graph in Spark Without SPARQL”, Database and Expert Systems Applications, DEXA 2020 International Workshops BIOKDD, IWCFS and MLKgraphs, Bratislava, Slovakia, September 14–17, 2020, Proceedings.

[23] "Find New Associations in Text",http://sparklingdataocean.com/2018/04/04/word2vec2graphInsights/

[24] "Word2Vec2Graph Model and Free Associations",http://sparklingdataocean.com/2017/12/24/word2vec2graphPsychoanalysis/

[25] Practical Deep Learning for Coders, https://course.fast.ai/ (2020).

[26] Jeremy Howard, Sylvain Gugger: Deep Learning for Coders with fast.ai and Py-Torch. O’Reilly Media, Inc. (2020).

[27] Time series/ sequential data study group, https://forums.fast.ai/t/time-series-sequential-data-studygroup/29686 (2019)

[28] "GoodTherapy: PsychPedia: Free Association", https://www.goodtherapy.org/blog/psychpedia/free-association-in-therapy (2019).

[29] "Word2Vec2Graph to Images to Deep Learning",http://sparklingdataocean.com/2019/03/16/word2vec2graph2CNN/

[30] “Practical Deep Learning applied to Time Series", https://github.com/oguiza

[31] “Motifs Findings in GraphFrames”, https://www.waitingforcode.com/apachespark-graphframes/motifs-finding-graphframes/read

[32] “Drawing graphs with dot”, https://www.ocf.berkeley.edu/~eek/index.html/tiny_examples/thinktank/src/gv1.7c/doc/dotguide.p df

[33] “Visual network analysis with Gephi”, https://medium.com/@EthnographicMachines/visualnetwork-analysis-with-gephi-d6241127a336

[34] “EEG Patterns by Deep Learning and Graph Mining”,http://sparklingdataocean.com/2020/08/19/brainGraphEeg/

[35] Something2vec,https://gist.github.com/nzw0301/333afc00bd508501268fa7bf40cafe4e(2016)

A COMPREHENSIVE ANALYSIS OF QUANTUM CLUSTERING : FINDING ALL THE POTENTIAL MINIMA

Aude Maignan1 and Tony Scott2

1Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France 2 Institut für Physikalische Chemie, RWTH-Aachen University, 52056 Aachen, Germany

ABSTRACT

Quantum clustering (QC), is a data clustering algorithm based on quantum mechanics which is accomplished by substituting each point in a given dataset with a Gaussian. The width of the Gaussian is a σ value, a hyper-parameter which can be manually defined and manipulated to suit the application. Numerical methods are used to find all the minima of the quantum potential as they correspond to cluster centers. Herein, we investigate the mathematical task of expressing and finding all the roots of the exponential polynomial corresponding to the minima of a two-dimensional quantum potential. This is an outstanding task because normally such expressions are impossible to solve analytically. However, we prove that if the points are all included in a square region of size σ, there is only one minimum. This bound is not only useful in the number of solutions to look for, by numerical means, it allows to to propose a new numerical approach "per block". This technique decreases the number of particles by approximating some groups of particles to weighted particles. These findings are not only useful to the quantum clustering problem but also for the exponential polynomials encountered in quantum chemistry, Solid-state Physics and other applications.

KEYWORDS

Data clustering, Quantum clustering, energy function, exponential polynomial, optimization.

Full Text: https://aircconline.com/ijdkp/V11N1/11121ijdkp03.pdf

REFERENCES

[1] M. Fertik, T Scott and T Dignan, US Patent No. 2013/0086075 A1, Appl. No. 13/252,697 - Ref. US9020952B2, (2013)

[2] D. Horn and A. Gottlieb, Phys. Rev. Lett. 88, 18702 (2002)

[3] T. C. Scott, M. Therani and X. M. Wang, Mathematics 5, 1-17 (2017)

[4] A. Ben-Hur, D. Horn, H. T. Siegelmann and V. Vapnik, J. Mach. Learn. Res. 2, 125-137 (2002)

[5] A. Messiah, Quantum Mechanics (Vol. I), English translation from French by G. M. Temmer, North Holland, John Wiley & Sons, Cf. chap. IV, section III. chap. 3, sec.12, 1966.

[6] A. Lüchow and T. C. Scott, J. Phys. B: At. Mol. Opt. Phys. 40, 851-867 (2007)

[7] A. Lüchow, R. Petz R and T. C. Scott, J. Chem. Phys. 126, 144110-144110 (2007) [8] T. C. Scott, A. Lüchow, D. Bressanini and J.D. Morgan III, Phys. Rev. A (Rapid Communications) 75, 060101 (2007)

[9] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, et al., Gaussian Inc., Wallingford CT (2009)

[10] T. C. Scott, I. P. Grant, M. B. Monagan and V. R. Saunders, Nucl. Instruments and Methods Phys. Research 389A, 117-120 (1997)

[11] T. C. Scott, I. P. Grant, M. B. Monagan and V.R. Saunders, MapleTech 4, 15-24 (1997)

[12] C. Gomez and T. C. Scott, Comput. Phys. Commun. 115, 548-562 (1998)

[13] Achatz, M., McCallum, S., & Weispfenning, V. (2008). Deciding polynomial-exponential problems. In D. Jeffrey (Ed.), ISSAC'08: Proceedings of the 21st International Symposium on Symbolic and Algebraic Computation 2008 (pp. 215-222). New York: Association for Computing Machinery. https://doi.org/10.1145/1390768.1390799

[14] A. Maignan, Solving One and Two-dimensional Exponential Polynomial Systems, ISSAC98, ACM press, pp 215-221.

[15] Scott McCallum, Volker Weispfenning, Deciding polynomial-transcendental problems, Journal of Symbolic Computation, Volume 47, Issue 1, 2012, Pages 16-31, ISSN 0747-7171, https://doi.org/10.1016/j.jsc.2011.08.004.

[16] J. F. Rodriguez-Nieva and M. S. Scheurer, Identifying topological order through unsupervised machine learning, Nature Physics, Nature, Physics 15, 790 (2019)

[17] Eran Lustig, Or Yair, Ronen Talmon, and Mordechai Segev, Identifying Topological Phase Transitions in Experiments Using Manifold Learning, Phys. Rev. Lett. 125, 127401 (2020)

[18] Jielin Wang, Wanzhou Zhang, Tian Hua and Tzu-Chieh Wei, Unsupervised learning of topological phase transitions using Calinski-Harabaz score, accepted by Physical Review Research, (2020)

[19] Shervan Fekri Ershad, Texture Classification Approach Based on Energy Variation IJMT 2, 52-55 (2012)

[20] Fan Decheng, Song Jon, Cholho Pang, Wang Dong, CholJin Won, Improved quantum clustering analysis based on the weighted distance and its application, Heliyon, Volume 4, Issue 11, 2018, e00984, ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2018.e00984.

[21] A. Maignan, On Symbolic-Numeric Solving of Sine-Polynomial Equations, Journal of Complexity, vol 16, Issue 1, (2000), pp. 274-285

[22] A. Maignan and T. C. Scott, SIGSAM 50, 45-60 (2016)

[23] https://proofwiki.org/wiki/Upper_Bound_of_Natural_Logarithm

[24] B. Ripley, Cambridge University Press , Cambridge, UK (1996)

[25] B. Ripley, Available online , http://www.stats.ox.ac.uk/pub/PRNN/ (accessed on 3 January 2017)

[26] L. Bernardin, P. Chin, P. DeMarco, K. O. Geddes, D. E. G. Hare, K. M. Heal, G. Labahn, J. P. May, J. McCarron, M. B. Monagan, D. Ohashi and S. M. Vorkoetter, MapleSoft , Toronto (2012)

[27] I. Mezo and A. Baricz, On the generalization of the Lambert W function, Transactions of the American Mathematical Society 369, 7917-7934 (2017).

[28] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, Advances in Computational

Mathematics 5, 329-359 (1996)

[29] T. C. Scott, G. J. Fee and J. Grotendorst, SIGSAM 47, 75-83 (2013)

[30] T. C. Scott, G. J. Fee, J. Grotendorst and W.Z. Zhang, SIGSAM 48, 42-56 (2014)

[31] T. C. Scott and A. Maignan, SIGSAM 50, 45-60 (2016)

[32] T. C. Scott, A. Dalgarno and J. D. Morgan III, Phys. Rev. Lett. 67, 1419-1422 (1991)

[33] T. C. Scott, J. F. Babb, A. Dalgarno and J. D. Morgan III, J. Chem. Phys. 99, 2841-2854 (1993)

[34] T. C. Scott, M. Aubert-Frécon and J. Grotendorst, Chem. Phys. 324, 323-338 (2006)

[35] T. C. Scott and R. B. Mann and R. E. Martinez, AAECC (Applicable Algebra in Engineering, Communication and Computing) 17, 41-47 (2006)

[36] P. S. Farrugia, R. B. Mann, and T. C. Scott. Class. Quantum Grav 24, 4647-4659 (2007)

[37] Thanh-Toan Pham, PhD Thesis in Electronics, University of Grenoble Alpes (ComUE) , (2017)

[38] Exoplanet.eu-Extrasolar Planets Encyclopedia, Available online , http://exoplanet.eu/ Retrieved 16 November 2015 (accessed on 2 January 2017)

[39] A. Maignan, T. C. Scott, Quantum Clustering Analysis: Minima of the Potential Energy Function, ISSN : 2231 - 5403 ,Vol. 10 – vol. NO: 19 - Issue: 19/12/2020.

[40] A. Gionis, H. Mannila, and P. Tsaparas, Clustering aggregation. ACM Transactions on Knowledge Discovery from Data (TKDD), 2007. 1(1): p. 1-30.

[41] L. Fu and E. Medico, FLAME, a novel fuzzy clustering method for the analysis.

[42] N.R. Draper and H. Smith, “Applied Regression Analysis”, 2nd ed., Wiley, New York, (1981).

[43] Forrest R. Miller, James W. Neill and Brian W. Sherfey, “Maximin Clusters from near-replicate Regression of Fit Tests”, Ann. Stat. 26, no. 4, pp. 1411-1433, (1998).

[44] Isak Gath and Dan Hoory, Fuzzy clustering of elliptic ring-shaped clusters, Pattern Recognition Letters", Vol. 16, 1995, p. 727-741, https://doi.org/10.1016/0167-8655(95)00030-K.

[45] http://cs.joensuu.fi/sipu/datasets/ [46] A. Ahmad and S. S. Khan, "Survey of State-of-the-Art Mixed Data Clustering Algorithms," in IEEE Access, vol. 7, pp. 31883-31902, 2019, doi: 10.1109/ACCESS.2019.2903568.

PARTITIONING WIDE AREA GRAPHS USING A SPACE FILLING CURVE

Cyprien Gottstein1 , Philippe Raipin Parvedy1 , Michel Hurfin2 , Thomas Hassan1 and Thierry Coupaye

1 1TGI-OLS-DIESE-LCP-DDSD, Orange Labs, Cesson-Sevigné, France 2Univ Rennes,INRIA, CNRS, IRISA, 35000RENNES, France

ABSTRACT

Graph structure is a very powerful tool to model system and represent their actual shape. For instance, modelling an infrastructure or social network naturally leads to graph. Yet, graphs can be very different from one another as they do not share the same properties (size, connectivity, communities, etc.) and building a system able to manage graphs should take into account this diversity. A big challenge concerning graph management is to design a system providing a scalable persistent storage and allowing efficient browsing. Mainly to study social graphs, the most recent developments in graph partitioning research often consider scale-free graphs. As we are interested in modelling connected objects and their context, we focus on partitioning geometric graphs. Consequently our strategy differs, we consider geometry as our main partitioning tool. In fact, we rely on Inverse Space-filling Partitioning, a technique which relies on a space filling curve to partition a graph and was previously applied to graphs essentially generated from Meshes. Furthermore, we extend Inverse Space-Filling Partitioning toward a new target we define as Wide Area Graphs. We provide an extended comparison with two state-of-the-art graph partitioning streaming strategies, namely LDG and FENNEL. We also propose customized metrics to better understand and identify the use cases for which the ISP partitioning solution is best suited. Experimentations show that in favourable contexts, edge-cuts can be drastically reduced, going from more 34% using FENNEL to less than 1% using ISP.

KEYWORDS

Graph, Partitioning, Graph partitioning, Geometric partitioning, Spatial, Geography, Geometric, Space Filling Curve, SFC, ISP

Full Text: https://aircconline.com/ijdkp/V11N1/11121ijdkp02.pdf

REFERENCES

[1] K. Andreev et H. Räcke, « Balanced Graph Partitioning », in Proceedings of the Sixteenth Annual ACM Symposium on Parallelism in Algorithms and Architectures, New York, NY, USA, 2004, p. 120–124, doi: 10.1145/1007912.1007931.

[2] « Recent Advances in Graph Partitioning | SpringerLink ». https://link.springer.com/chapter/10.1007/978-3-319-49487-6_4 (consulté le oct. 07, 2020).

[3] F. Payan, C. Roudet, et B. Sauvage, « Semi-Regular Triangle Remeshing: A Comprehensive Study », Comput. Graph. Forum, vol. 34, no 1, p. 86-102, 2015, doi: 10.1111/cgf.12461.

[4] J. R. Pilkington et S. B. Baden, « Partitioning with Spacefilling Curves », 1994.

[5] C. Tsourakakis, C. Gkantsidis, B. Radunovic, et M. Vojnovic, « FENNEL: streaming graph partitioning for massive scale graphs », in Proceedings of the 7th ACM international conference on Web search and data mining - WSDM ’14, New York, New York, USA, 2014, p. 333-342, doi: 10.1145/2556195.2556213.

[6] G. Karypis et V. Kumar, « Multilevelk-way Partitioning Scheme for Irregular Graphs », J. Parallel Distrib. Comput., vol. 48, no 1, p. 96-129, janv. 1998, doi: 10.1006/jpdc.1997.1404.

[7] « Partitioning of unstructured problems for parallel processing - ScienceDirect ». https://www.sciencedirect.com/science/article/abs/pii/095605219190014V (consulté le janv. 20, 2020).

[8] J. R. Gilbert, G. L. Miller, et S.-Hua. Teng, « Geometric Mesh Partitioning: Implementation and Experiments », SIAM J. Sci. Comput., vol. 19, no 6, p. 2091-2110, nov. 1998, doi: 10.1137/S1064827594275339.

[9] K. Schloegel, G. Karypis, et V. Kumar, « Graph partitioning for high-performance scientific simulations », in Sourcebook of parallel computing, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003, p. 491–541.

[10] A. Akdogan, « Partitioning, Indexing and Querying Spatial Data on Cloud », p. 80.

[11] S. Agarwal, J. Dunagan, N. Jain, S. Saroiu, A. Wolman, et H. Bhogan, « Volley: Automated Data Placement for Geo-Distributed Cloud Services », p. 16.

[12] D. Delling, A. V. Goldberg, I. Razenshteyn, et R. F. Werneck, « Graph Partitioning with Natural Cuts », in 2011 IEEE International Parallel Distributed Processing Symposium, mai 2011, p. 1135-1146, doi: 10.1109/IPDPS.2011.108.

[13] I. Stanton et G. Kliot, « Streaming graph partitioning for large distributed graphs », in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’12, Beijing, China, 2012, p. 1222, doi: 10.1145/2339530.2339722.

[14] D. Hilbert, « Über die stetige Abbildung einer Linie auf ein Flächenstück », in Dritter Band: Analysis • Grundlagen der Mathematik • Physik Verschiedenes: Nebst Einer Lebensgeschichte, Berlin, Heidelberg: Springer Berlin Heidelberg, 1935, p. 1–2.

[15] M. Knoll et T. Weis, « Optimizing Locality for Self-organizing Context-Based Systems », in SelfOrganizing Systems, 2006, p. 62-73.

[16] B. M. Waxman, « Routing of multipoint connections », IEEE J. Sel. Areas Commun., vol. 6, no 9, p. 1617-1622, déc. 1988, doi: 10.1109/49.12889.

[17] M. D. Penrose, « Connectivity of soft random geometric graphs », Ann. Appl. Probab., vol. 26, no 2, p. 986-1028, avr. 2016, doi: 10.1214/15-AAP1110.

[18] R. Albert et A.-L. Barabasi, « Statistical mechanics of complex networks », Rev. Mod. Phys., vol. 74,

no 1, p. 47-97, janv. 2002, doi: 10.1103/RevModPhys.74.47.

[19] A. Guttman, « R-trees: a dynamic index structure for spatial searching », in Proceedings of the 1984 ACM SIGMOD international conference on Management of data, New York, NY, USA, juin 1984, p. 47–57, doi: 10.1145/602259.602266.Technology (IJCET), 5(10), 01- 10.

APPLY MACHINE LEARNING METHODS TO PREDICT FAILURE OF GLAUCOMA DRAINAGE

Paul Morrison1 , Maxwell Dixon2 , Arsham Sheybani2 and Bahareh Rahmani1

1Fontbonne University, Mathematics and Computer Science Department, St. Louis, MO
2Washington University, Department of Ophthalmology and Visual Sciences, St. Louis, MO

ABSTRACT

The purpose of this retrospective study is to measure machine learning models' ability to predict glaucoma drainage device failure based on demographic information and preoperative measurements. The medical records of 165 patients were used. Potential predictors included the patients' race, age, sex, preoperative intraocular pressure (IOP), preoperative visual acuity, number of IOP-lowering medications, and number and type of previous ophthalmic surgeries. Failure was defined as final IOP greater than 18 mm Hg, reduction in intraocular pressure less than 20% from baseline, or need for reoperation unrelated to normal implant maintenance. Five classifiers were compared: logistic regression, artificial neural network, random forest, decision tree, and support vector machine. Recursive feature elimination was used to shrink the number of predictors and grid search was used to choose hyperparameters. To prevent leakage, nested cross-validation was used throughout. With a small amount of data, the best classfier was logistic regression, but with more data, the best classifier was the random forest.

Full Text: https://aircconline.com/ijdkp/V11N1/11121ijdkp01.pdf

REFERENCES

[1] D. L. Budenz et al., "Five-Year Treatment Outcomes in the Ahmed Baerveldt Comparison Study," Ophthalmology, vol. 122, no. 2, pp. 308–316, Feb. 2015, doi: 10.1016/j.ophtha.2014.08.043.

[2] A. Achiron et al., "Predicting Refractive Surgery Outcome: Machine Learning Approach With Big Data," J Refract Surg, vol. 33, no. 9, pp. 592–597, Sep. 2017, doi: 10.3928/1081597X-20170616-03.

[3] M. Rohm et al., "Predicting Visual Acuity by Using Machine Learning in Patients Treated for Neovascular Age-Related Macular Degeneration," Ophthalmology, vol. 125, no. 7, pp. 1028–1036, Jul. 2018, doi: 10.1016/j.ophtha.2017.12.034.

[4] M. A. Valdes-Mas, J. D. Martin, M. J. Ruperez, C. Peris, and C. Monserrat, "Machine learning for predicting astigmatism in patients with keratoconus after intracorneal ring implantation," in IEEEEMBS International Conference on Biomedical and Health Informatics (BHI), Valencia, Spain, Jun. 2014, pp. 756–759, doi: 10.1109/BHI.2014.6864474.

[5] S.-F. Mohammadi et al., "Using artificial intelligence to predict the risk for posterior capsule opacification after phacoemulsification:," Journal of Cataract & Refractive Surgery, vol. 38, no. 3, pp. 403–408, Mar. 2012, doi: 10.1016/j.jcrs.2011.09.036.

[6] M. Gupta, P. Gupta, P. K. Vaddavalli, and A. Fatima, "Predicting Post-operative Visual Acuity for LASIK

Surgeries," in Advances in Knowledge Discovery and Data Mining, vol. 9651, J. Bailey, L. Khan, T. Washio, G. Dobbie, J. Z. Huang, and R. Wang, Eds. Cham: Springer International Publishing, 2016, pp. 489–501.

[7] R. Koprowski, M. Lanza, and C. Irregolare, "Corneal power evaluation after myopic corneal refractive surgery using artificial neural networks," BioMed EngOnLine, vol. 15, no. 1, p. 121, Dec. 2016, doi: 10.1186/s12938-016-0243-5.

[8] R. P. McNabb, S. Farsiu, S. S. Stinnett, J. A. Izatt, and A. N. Kuo, "Optical Coherence Tomography Accurately Measures Corneal Power Change from Laser Refractive Surgery," Ophthalmology, vol. 122, no. 4, pp. 677–686, Apr. 2015, doi: 10.1016/j.ophtha.2014.10.003.

[9] C. Bowd et al., "Predicting Glaucomatous Progression in Glaucoma Suspect Eyes Using Relevance Vector Machine Classifiers for Combined Structural and Functional Measurements," Invest. Ophthalmol. Vis. Sci., vol. 53, no. 4, p. 2382, Apr. 2012, doi: 10.1167/iovs.11-7951.

[10] J. Lee, Y. K. Kim, J. W. Jeoung, A. Ha, Y. W. Kim, and K. H. Park, "Machine learning classifiers-based prediction of normal-tension glaucoma progression in young myopic patients," Jpn J Ophthalmol, vol. 64, no. 1, pp. 68–76, Jan. 2020, doi: 10.1007/s10384-019-00706-2.

[11] S. L. Baxter, C. Marks, T.-T. Kuo, L. Ohno-Machado, and R. N. Weinreb, "Machine LearningBased Predictive Modeling of Surgical Intervention in Glaucoma Using Systemic Data FromElectronic Health Records," American Journal of Ophthalmology, vol. 208, pp. 30–40, Dec. 2019, doi: 10.1016/j.ajo.2019.07.005.

[12] D. Krstajic, L. J. Buturovic, D. E. Leahy, and S. Thomas, "Cross-validation pitfalls when selecting and assessing regression and classification models," J Cheminform, vol. 6, no. 1, p. 10, Dec. 2014, doi: 10.1186/1758-2946-6-10.

[13] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," arXiv:1201.0490 [cs], Jun. 2018, Accessed: Nov. 22, 2020. [Online]. Available: http://arxiv.org/abs/1201.0490.

[14] M. Kuhn, "Building Predictive Models in R Using the caret Package," J. Stat. Soft., vol. 28, no. 5, 2008, doi: 10.18637/jss.v028.i05.

[15] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to data mining, 1st ed. Boston: Pearson Addison Wesley, 2006.

[16] "The advanced glaucoma intervention study (AGIS)*113. Comparison of treatment outcomes within race: 10-year results," Ophthalmology, vol. 111, no. 4, pp. 651–664, Apr. 2004, doi: 10.1016/j.ophtha.2003.09.025.

[17] D. Broadway, I. Grierson, and R. Hitchings, "Racial differences in the results of glaucoma filtration surgery: are racial differences in the conjunctival cell profile important?," British Journal of Ophthalmology, vol. 78, no. 6, pp. 466–475, Jun. 1994, doi: 10.1136/bjo.78.6.466.

[18] K. Inoue, "Managing adverse effects of glaucoma medications," OPTH, p. 903, May 2014, doi: 10.2147/OPTH.S44708.

PETROCHEMICAL PRODUCTION BIG DATA AND ITS FOUR TYPICAL APPLICATION PARADIGMS

Hu Shaolin* , Zhang Qinghua, Su Naiquan, and Li Xiwu

Guangdong University of Petrochemical Technology, Maoming, Guangdong, China

ABSTRACT

In recent years, the big data has attracted more and more attention. It can bring us more information and broader perspective to analyse and deal with problems than the conventional situation. However, so far, there is no widely acceptable and measurable definition for the term “big data”. For example, what significant features a data set needs to have can be called big data, and how large a data set is can be called big data, and so on. Although the "5V" description widely used in textbooks has been tried to solve the above problems in many big data literatures, "5V" still has significant shortcomings and limitations, and is not suitable for completely describing big data problems in practical fields such as industrial production. Therefore, this paper creatively puts forward the new concept of data cloud and the data cloud-based "3M" descriptive definition of big data, which refers to a wide range of data sources (Multisource), ultra-high dimensions (Multi-dimensional) and a long enough time span (Multi-spatiotemporal). Based on the 3M description of big data, this paper sets up four typical application paradigms for the production big data, analyses the typical application of four paradigms of big data, and lays the foundation for applications of big data from petrochemical industry.

KEYWORDS

Big Data, Paradigms, Industrial Big Data.

Full Text: https://aircconline.com/ijdkp/V11N4/11421ijdkp02.pdf

REFERENCES

[1] Li Ning(2019).Artificial Intelligence Paradigm in Big Data Era. China Computer & Communication, 8:104-105.

[2] Jarosław W, Jarosław J, Paweł Z(2019), Generalised framework for Multi-criteria Method Selection. Omega,86:107–124.

[3] Manyika J, Chui M, Brown B, Bughin J.,et al(2017). Big data: The Next Frontier for Innovation, Competition, and productivity. McKinsey Global Institute

[4] Abbass H. A, Leu G,Merrick K(2016).A Review of Theoretical and Practical Challenges of Trusted Autonomy in Big Data. Theoretical Foundations for Big Data Applications: Challenges and Opportunities.

[5] Ammu N, Irfanuddin M. Big data challenges. International Journal of Advanced Trends in Computer Science and Engineering, 2013,2(1), 613-615

[6] Rabhi L, Falin N, Afraites A,et al(2019).Procedia Computer Science,16th International Conf. on Mobile Systems and Pervasive Computing, v155, pp:599-605.

[7] Mukherjee S, Shaw R(2016). Big Data-Concepts, Applications, Challenges and Future Scope. Wikipedia, https://en.wikipedia.org/wiki/Google_Cloud_Platform

[8] Tanya Garg;Surbhi Khullar(2020). Big Data Analytics: Applications, Challenges & Future Directions. 8th International Confer on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). pp:923-928,4-5 June 2020, Noida, India.

[9] Nojood Aljehane(2020).Big Data Analytics: Challenges and Opportunities. International Confer on Computing and Information Technology (ICCIT-1441), pp:1-4, 9-10 Sept. 2020, Tabuk, Saudi Arabia.

[10] Ma Z, Yang L, Zhang, Q(2021). Support Multimode Tensor Machine for Multiple Classification on Industrial Big Data. IEEE Transactions on Industrial Informatics,17(5):3382-3390.

[11] Sakineti S; Prabhu C (2018). Protagonist of Big Data and Predictive Analytics using Data Analytics. International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), pp:276-279,21-22 Dec. 2018, Belgaum,India.

[12] Zhou X, Hu Y, Liang W,et al(2021). Variational LSTM Enhanced Anomaly Detection for Industrial Big Data. IEEE Transactions on Industrial Informatics,17(5): 3469-3477.

[13] Sfaxi L,Ben A,Mohamed M(2020), DECIDE: An Agile Event-and-Data Driven Design Methodology for Decisional Big Data Projects. Knowledge Engineering, volume 130, DOI: 10.1016/j.datak.2020.101862.

[14] Battistelli G, Tesi, P(2021). Classification for Dynamical Systems: Model-Based and Data-Driven Approaches.IEEE Transactions on Automatic Control, 66(4):1741-1748.

Referring Expressions with Rational Speech Act Framework: A Probabilistic Approach

Hieu Le[1] , Taufiq Daryanto[2] , Fabian Zhafransyah[2] , Derry Wijaya[1] , Elizabeth Coppock[1] , Sang Chin[1]

[1] Boston University , Boston, MA, USA
[2] Institut Teknologi Bandung, Bandung, Indonesia

ABSTRACT

This paper focuses on a referring expression generation (REG) task in which the aim is to pick out an object in a complex visual scene. One common theoretical approach to this problem is to model the task as a two-agent cooperative scheme in which a 'speaker' agent would generate the expression that best describes a targeted area and a 'listener' agent would identify the target. Several recent REG systems have used deep learning approaches to represent the speaker/listener agents. The Rational Speech Act framework (RSA), a Bayesian approach to pragmatics that can predict human linguistic behavior quite accurately, has been shown to generate high quality and explainable expressions on toy datasets involving simple visual scenes. Its application to large scale problems, however, remains largely unexplored. This paper applies a combination of the probabilistic RSA framework and deep learning approaches to larger datasets involving complex visual scenes in a multi-step process with the aim of generating better-explained expressions. We carry out experiments on the RefCOCO and RefCOCO+ datasets and compare our approach with other endto-end deep learning approaches as well as a variation of RSA to highlight our key contribution. Experimental results show that while achieving lower accuracy than SOTA deep learning methods, our approach outperforms similar RSA approach in human comprehension and has an advantage over end-to-end deep learning under limited data scenario. Lastly, we provide a detailed analysis on the expression generation process with concrete examples, thus providing a systematic view on error types and deficiencies in the generation process and identifying possible areas for future improvements.

Full Text: https://aircconline.com/ijdkp/V12N3/12322ijdkp01.pdf

REFERENCES

[1] W. Monroe and C. Potts, "Learning in the Rational Speech Acts model," CoRR, vol. abs/1510.06807, 2015.

[2] S. Kazemzadeh, V. Ordonez, M. Matten, and T. Berg, "ReferItGame: Referring to objects in photographs of natural scenes," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), (Doha, Qatar), pp. 787–798, ACL, Oct. 2014.

[3] M. C. Frank and N. D. Goodman, "Predicting pragmatic reasoning in language games," Science, vol. 336, no. 6084, p. 998, 2012.

[4] J. Degen, R. X. D. Hawkins, C. Graf, E. Kreiss, and N. D. Goodman, "When redundancy is rational: A Bayesian approach to 'overinformative' referring expressions," CoRR, vol. abs/1903.08237, 2019.

[5] J. Andreas and D. Klein, "Reasoning about pragmatics with neural listeners and speakers," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, (Austin,

Texas), pp. 1173–1182, ACL, Nov. 2016.

[6] R. Cohn-Gordon, N. D. Goodman, and C. Potts, “Pragmatically informative image captioning with character-level reference,” CoRR, vol. abs/1804.05417, 2018

. [7] J. Yang, J. Lu, S. Lee, D. Batra, and D. Parikh, “Graph R-CNN for scene graph generation,” CoRR, vol. abs/1808.00191, 2018.

[8] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, “Detectron2.” https: //github.com/facebookresearch/detectron2, 2019.

[9] L. Yu, P. Poirson, S. Yang, A. C. Berg, and T. L. Berg, “Modeling context in referring expressions,” CoRR, vol. abs/1608.00272, 2016.

[10] G. Scontras, M. H. Tessler, and M. Franke, “Probabilistic language understanding: An introduction to the rational speech act framework,” 2018.

[11] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in Advances in Neural Information Processing Systems (NIPS), 2015.

[12] W. Abdulla, “Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow.” https://github.com/matterport/Mask_RCNN, 2017.

[13] R. Luo and G. Shakhnarovich, “Comprehension-guided referring expressions,” CoRR, vol. abs/1701.03439, 2017. [14] D. Lassiter and N. Goodman, “Adjectival vagueness in a Bayesian model of interpretation,” Synthese, vol. 10, pp. 3801–3836, 2017.

[15] D. Edgington, “The philosophical problem of vagueness,” Legal Theory, vol. 7, pp. 371–378, 2001.

[16] L. Yu, H. Tan, M. Bansal, and T. L. Berg, “A joint speaker-listener-reinforcer model for referring expressions,” 2017.

[17] J. Kim, H. Ko, and J. Wu, “CoNAN: A complementary neighboring-based attention network for referring expression generation,” in Proceedings of the 28th International Conference on Computational Linguistics, (Barcelona, Spain (Online)), pp. 1952– 1962, International Committee on Computational Linguistics, Dec. 2020.

[18] P. Joshi, S. Santy, A. Budhiraja, K. Bali, and M. Choudhury, “The state and fate of linguistic diversity and inclusion in the NLP world,” in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, (Online), pp. 6282–6293, Association for Computational Linguistics, July 2020.