IJCAI–19 Tutorial Proposal - Non-IID Learning of Complex Data and Behaviors

Longbing Cao[1] , **Trong Dinh Thac Do**[1] , **Chengzhang Zhu**[1]

[1]Advanced Analytics Institute - University of Technology Sydney

{Longbing.Cao, TrongDinhThac.Do}@uts.edu.au, kevin.zhu.china@gmail.com

1 Title

Non-IID Learning of Complex Data and Behaviors.

2 Brief Summary: A two-sentence description of the tutorial

Non-IID learning of complex data and behaviors refers to the methodologies, algorithms and practical tools for representing, modeling, analyzing and understanding non-IID (not independent and identically distributed, also non-i.i.d.) data through representing, learning and synthesizing the intrinsic heterogeneities and coupling relationships (non-IIDness for short). This tutorial introduces the methodologies on the representation of non-IIDness of complex data and behaviors and presents the state-of-the-art representation, statistical learning, metric learning, and similarity learning theories and task-oriented learning algorithms by incorporating non-IIDness into a variety of learning problems, including classification, clustering, ensemble clustering, outlier detection, feature selection, recommender systems, and text mining.

3 Abstract: A two-paragraph description of the tutorial

Learning of complex data and behaviors is increasingly becoming a major challenge and opportunity for big business and innovative learning theories and tools. Some of the most critical challenges of learning of complex data and behaviors are the uncovering of the explicit and implicit coupling relationships embedded in mixed heterogeneous data from single/multiple sources. The couplings and heterogeneities of the non-IIDness aspects form the essence of complex data and behaviors and most real-world applications, namely the data is non-IID.

Most of classic theoretical systems and tools in statistics, data mining, database, knowledge management and machine learning assume the independence and identical distribution of underlying objects, features and values. Such theories and tools may lead to misleading or incorrect understanding of real-life data complexities. Non-IID learning of complex data and behaviors is a foundational theoretical problem in AI and data science, which considers the couplings and heterogeneities between entities, properties, interactions and contexts. In this tutorial, we present a comprehensive tutorial of the non-IID learning. We begin with the limitation of IID learning in handling complex data and behaviors and introduce abstract learning model and representation for non-IID learning, and then present frameworks and algorithms for non-IID metric learning, classification, clustering, ensemble clustering, outlier detection, feature selection, recommender systems, and text mining, and finally discuss open challenges and prospects.

4 Proposed Length

The proposed length of the tutorial is $1/2$ day (two 1h 45mins slots).

5 Outline: A detailed, point-form outline of the tutorial

The tutorial includes the following contents:

- The main challenges in complex data and behaviors: a brief introduction of complex data and behaviors and the essential theoretical challenges in learning complex data and behaviors;

- What is IID learning: summarize the classic IID learning theories by several examples such as clustering and recommender system theories;

- Limitations of IID learning: summarize the main limitations of the classic IID learning methods, and introduce why they do not work well in handling complex data and behaviors;

- Related work of non-IID learning: introduce the related work of non-IID learning to address the gaps and issues in classic IID and non-IID data learning theories and models;

- Non-IID characteristics and nature: discuss the characteristics of non-IID data, especially the coupling relationships and heterogeneity in complex data and behaviors;

- Abstract non-IID learning model and representation: discuss an abstract representation model of non-IIDness by involving the heterogeneity and couplings between entities, properties, contexts and interactions etc.;

- Non-IID clustering: new similarity metrics, algorithms and case studies will be presented to introduce non-IIDness into clustering;

- Non-IID ensemble clustering: new similarity metrics, algorithms and case studies will be presented to cater for non-IIDness in ensemble clustering;

- Non-IID outlier detection: new methods will be introduced to incorporate non-IIDness into outlier detection and outlying feature selection;

- Non-IID classification: similarity metrics and algorithms will be introduced to incorporate non-IIDness into classification;

- Non-IID recommender systems: non-IID recommender systems including the framework, algorithms and experiments will be introduced;

- Term coupling-based document analysis: similarity metrics, algorithms and case study of incorporating non-IID learning into document analysis will be discussed;

- Hierarchical structure and couplings in non-IID learning: We summarize the case studies and conclude the hierarchical structure and couplings in non-IID learning as a foundation for non-IID learning;

- Challenges and prospects of non-IID information processing and management: open issues and opportunities will be discussed for incorporating non-IIDness into information processing and management, including database management, knowledge management, information retrieval, recommender system, data mining and machine learning to develop corresponding methodologies, systems and algorithms for real-world big data applications.

6 Description of Intended Audience: A characterization of the potential target audience

Non-IID data learning is a fundamental issue in data-driven AI. Any audience who may be interested in data-driven discovery and AI theories and systems, for instance knowledge representation and management, machine learning, data mining, image processing, computer vision, recommender system, online network analysis, social media analysis, and text mining, would find it very inspiring and valuable in attending this tutorial. In this tutorial, the attendant will not only learn the basic knowledge of non-IID learning and Big Data, but also can benefit from the detailed theories and algorithms for solving the IID learning challenges in the above areas and domains.

While no specific knowledge is required from the audience, people who are familiar with the above mentioned areas will find it more beneficial in understanding the algorithms and case studies to be introduced in this tutorial.

7 Motivation: A description of why the tutorial topic would be of interest

There has been increasing attention from both the AI academic and industry communities, witnessed by the development and adoption of various autonomous systems or devices, e.g., autonomous cars, drones and personal assistants. Data science plays a major role in this generation of AI and machine intelligence enhancement, which involves such techniques and methods as machine learning and data mining, knowledge representation and reasoning, natural language processing, reinforcement learning.

Learning explicit and especially implicit non-IIDness is fundamental, yet challenging, for an in-depth understanding of data with weak and/or unclear structures, distributions, relationships, and semantics. In many cases, locally explicit but globally implicitly (or vice versa) non-IIDness are presented in a range of forms, structures, and layers and on diverse entities. Often, individual learners cannot tell the whole story due to their inability to identify such complex non-IIDness. Effectively learning the widespread, various, explicit and implicit non-IIDness is thus crucial for obtaining the truth and a complete picture of the underlying data characteristics and complexities. This is expected to be of great interest to a wide proportion of the IJCAI audience, e.g., researchers from the communities of machine learning, data mining, knowledge representation and reasoning, computer vision, image processing, natural language processing, document analysis, and recommender systems.

8 Background Information: A brief resume of the presenters

8.1 Names, affiliations, homepages and contact details of all presenters

Professor Longbing Cao
  Affiliation: Advanced Analytics Institute, Faculty of Engineering and IT, University of Technology Sydney, Australia
  Address: 2 Blackfriars St, Chippendale, 2008, Australia
  Tel: +61-2-9514-4477
  Fax: +61-2-9514-1807
  Homepage: cao.datasciences.org/
  Email: longbing.cao@uts.edu.au

Trong Dinh Thac Do
  Affiliation: Advanced Analytics Institute, Faculty of Engineering and IT, University of Technology Sydney, Australia
  Address: 2 Blackfriars St, Chippendale, 2008, Australia
  Homepage: sites.google.com/site/trongdinhthacdo/home
  Email: trongdinhthac.do@uts.edu.au

Chengzhang Zhu
  Affiliation: Advanced Analytics Institute, Faculty of Engineering and IT, University of Technology Sydney, Australia
  Address: 2 Blackfriars St, Chippendale, 2008, Australia
  Tel: +61-4-5208-0620
  Email: kevin.zhu.china@gmail.com

8.2 Short biographies of all presenters

Longbing Cao is a professor of information technology at the University of Technology Sydney (UTS), Australia. He holds a Ph.D. in Pattern Recognition and Intelligent Systems in Chinese Academy of Sciences and another Ph.D. in Computing Science at UTS. He has published more than 300 publications, four monographs, and four edited books in recent 15 years. He has been working on data science and analytics research, education, development, and enterprise applications since he was a CTO and then joined UTS. Motivated by real-world significant and common challenges, he has been leading the team to develop theories, tools and applications for new areas including non-IID learning, actionable knowledge discovery, behavior informatics, and complex intelligent systems, in addition to issues generally concerned in artificial intelligence, knowledge discovery, machine learning, and their enterprise applications. In data science and analytics, he initiated the Data Science and Knowledge Discovery lab at UTS in 2007, the Advanced Analytics Institute in 2011, the degrees Master of Analytics (Research) and PhD in Analytics in 2011 which are recognized as the world first degrees in data science, the IEEE Task Force on Data Science and Advanced Analytics (DSAA) and IEEE Task Force on Behavior, Economic and Soci-cultural Computing in 2013, the IEEE Conference on Data Science and Advanced Analytics (DSAA), the ACM SIGKDD Australia and New Zealand Chapter in 2014, and the International Journal of Data Science and Analytics with Springer in 2015. He served as program and general chairs of conferences such as KDD2015. In enterprise data science innovation, his team has successfully delivered many large projects for government and business organizations in over 10 domains including finance/capital markets, banking, health and car insurance, health, telco, recommendation, online business, education, and the public sector including ATO, DFS, DHS, DIBP and IP Australia, resulting in billions of dollar savings and mentions in government, industry, media and OECD reports. In 2013, AAI was the only organization specially mentioned in the Governments first big data paper: Big Data Strategy Issues Paper. He has been delivered invited and keynote speeches to over 20 conferences, guest lectures, and seminars to many universities, and tutorials to conferences including AAAI, IJCAI, and KDD.

Trong Dinh Thac Do is currently a Research Associate in the Advanced Analytics Institute (AAI) at the University of Technology Sydney (UTS), where he also obtained his Ph.D. degree. Before joining AAI-UTS, he received his Master of Philosophy degree in Information Technology in a joint program from the Grenoble Institute of Technology (INPG) and Joseph Fourier University (UJF), France. His research interests include machine learning, statistical models and Bayesian non-parametrics. He has been working in applied machine learning applications since he was a CTO and software engineer. He has published several papers in refereed conferences and journals, such as NIPS (new name NeurIPS), AAAI, IJCAI, ICDM, and International Journal of Knowledge and Information Systems (KAIS). He has served the community as program committee member or reviewer of IJCAI 2019, IJCAI 2018, AAAI 2019, ICDM 2018, ICDM 2017, PAKDD 2019,

PAKDD 2018, and reviewer of IEEE Transactions on Knowledge and Data Engineering (TKDE), IEEE Intelligent Systems, International Journal of Knowledge and Information Systems (KAIS), and International Journal of Data Science and Analytics (JDSA).

Chengzhang Zhu is currently a Ph.D. student in the Advanced Analytics Institute (AAI) at the University of Technology Sydney (UTS). His research interests include metric learning, non-IID learning, and data representation, in addition to general interest in data science especially machine learning. He has published several papers in refereed conferences and journals, such as IEEE Transactions on Knowledge and Data Engineering (TKDE), IJCAI, and ACM Multimedia Conference on Multimedia Conference (MM). He has served the community as program committee member of AAAI 2019, AAAI 2018, AAAI 2017, IJCAI 2019, IJCAI 2018, IJCAI 2017, CIKM 2019, CIKM 2018, PAKDD 2019, PAKDD 2018, PAKDD 2017, PAKDD 2016, and reviewer of IEEE Transactions on Systems, Man and Cybernetics: System, Information Science, IEEE Intelligent Systems, IEEE Access, and International Journal of Data Science and Analytics (JDSA).

8.3 Background in the tutorial area: A list of publications/presentations

[1] L. Cao. Combined mining: Analyzing object and pattern relations for discovering and constructing complex yet actionable patterns. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 3(2):140–155, 2013.

[2] L. Cao. Non-iidness learning in behavioral and social data. *Comput. J.*, 57(9):1358–1370, 2014.

[3] L. Cao. Coupling learning of complex interactions. *Inf. Process. Manage.*, 51(2):167–186, 2015.

[4] L. Cao. *Metasynthetic Computing and Engineering of Complex Systems*. Advanced Information and Knowledge Processing. Springer, 2015.

[5] L. Cao. Data science and analytics: a new era. *I. J. Data Science and Analytics*, 1(1):1–2, 2016.

[6] L. Cao. Data science: Nature and pitfalls. *IEEE Intelligent Systems*, 31(5):66–75, 2016.

[7] L. Cao. Non-iid recommender systems: A review and framework of recommendation paradigm shifting. *Engineering*, 2(2):212–224, 2016.

[8] L. Cao. Data science: A comprehensive overview. *ACM Comput. Surv.*, 50(3):43:1–43:42, 2017.

[9] L. Cao. Data science: challenges and directions. *Commun. ACM*, 60(8):59–68, 2017.

[10] L. Cao. *Data Science Thinking: The Next Scientific, Technological and Economic Revolution*. Data Analytics. Springer International Publishing, 2018.

[11] L. Cao and R. Dai. *Open Complex Intelligent Systems: Fundamentals, Concepts, Analysis, Design and Implementation (in Chinese)*. Posts & Telecom Press, 2008.

[12] L. Cao, J. Li, C. Wang, and P. S. Yu. Efficient selection of globally optimal rules on large imbalanced data based on rule coverage relationship analysis. In *Proceedings of the 13th SIAM International Conference on Data Mining, May 2-4, 2013. Austin, Texas, USA.*, pages 216–224, 2013.

[13] L. Cao, Y. Ou, and P. S. Yu. Coupled behavior analysis with applications. *IEEE Trans. Knowl. Data Eng.*, 24(8):1378–1392, 2012.

[14] L. Cao, Y. Ou, P. S. Yu, and G. Wei. Detecting abnormal coupled sequences and sequence changes in group-based manipulative trading behaviors. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010*, pages 85–94, 2010.

[15] L. Cao and P. Yu. Non-iid recommendation theories and systems. *IEEE Intelligent Systems*, 31(2):81–84, 2016.

[16] L. Cao and P. S. Yu. *Behavior Computing - Modeling, Analysis, Mining and Decision*. Springer, 2012.

[17] L. Cao, P. S. Yu, C. Zhang, and H. Z. (Eds.), editors. *Data Mining for Business Applications*. Springer, 2008.

[18] L. Cao, P. S. Yu, C. Zhang, and Y. Zhao. *Domain Driven Data Mining*. Springer, 2010.

[19] L. Cao, H. Zhang, Y. Zhao, D. Luo, and C. Zhang. Combined mining: Discovering informative knowledge in complex data. *IEEE Trans. Systems, Man, and Cybernetics, Part B*, 41(3):699–712, 2011.

[20] W. Cao, L. Cao, and Y. Song. Coupled market behavior based financial crisis detection. In *The 2013 International Joint Conference on Neural Networks, IJCNN 2013, Dallas, TX, USA, August 4-9, 2013*, pages 1–8, 2013.

[21] W. Cao, L. Hu, and L. Cao. Deep modeling complex couplings within financial markets. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 2518–2524, 2015.

[22] Q. Chen, L. Hu, J. Xu, W. Liu, and L. Cao. Document similarity analysis via involving both explicit and implicit semantic couplings. In *2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015, Campus des Cordeliers, Paris, France, October 19-21, 2015*, pages 1–10, 2015.

[23] X. Cheng, D. Miao, C. Wang, and L. Cao. Coupled term-term relation analysis for document clustering. In *The 2013 International Joint Conference on Neural Networks, IJCNN 2013, Dallas, TX, USA, August 4-9, 2013*, pages 1–8, 2013.

[24] T. D. T. Do and L. Cao. Coupled poisson factorization integrated with user/item metadata for modeling popular and sparse ratings in scalable recommendation. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the*

8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pages 2918–2925, 2018.

[25] T. D. T. Do and L. Cao. Gamma-poisson dynamic matrix factorization embedded with metadata influence. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada.*, pages 5829–5840, 2018.

[26] T. D. T. Do and L. Cao. Metadata-dependent infinite poisson factorization for efficiently modelling sparse and large matrices in recommendation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 5010–5016, 2018.

[27] L. C. (Ed.), editor. *Data Mining and Multi-agent Integration*. Springer, 2009.

[28] X. Fan, L. Cao, and R. Y. D. Xu. Dynamic infinite mixed-membership stochastic blockmodel. *IEEE Trans. Neural Netw. Learning Syst.*, 26(9):2072–2085, 2015.

[29] X. Fan, R. Y. D. Xu, and L. Cao. Copula mixed-membership stochastic blockmodel. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pages 1462–1468, 2016.

[30] X. Fan, R. Y. D. Xu, L. Cao, and Y. Song. Learning nonparametric relational models by conjugately incorporating node information in a network. *IEEE Trans. Cybernetics*, 47(3):589–599, 2017.

[31] S. Hao, C. Shi, Z. Niu, and L. Cao. Concept coupling learning for improving concept lattice-based document retrieval. *Eng. Appl. of AI*, 69:65–75, 2018.

[32] Y. Huang, L. Cao, J. Zhang, L. Pan, and Y. Liu. Exploring feature coupling and model coupling for image source identification. *IEEE Trans. Information Forensics and Security*, 13(12):3108–3121, 2018.

[33] S. Jian, L. Cao, K. Lu, and H. Gao. Unsupervised coupled metric similarity for non-iid categorical data. *IEEE Trans. Knowl. Data Eng.*, 30(9):1810–1823, 2018.

[34] S. Jian, L. Cao, G. Pang, K. Lu, and H. Gao. Embedding-based representation of categorical data by hierarchical value coupling learning. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 1937–1943, 2017.

[35] S. Jian, L. Hu, L. Cao, and K. Lu. Metric-based auto-instructor for learning mixed data representation. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 3318–3325, 2018.

[36] S. Jian, G. Pang, L. Cao, K. Lu, and H. Gao. Cure: Flexible categorical data representation by hierarchical

coupling learning. *IEEE Transactions on Knowledge and Data Engineering*, 2018.

[37] X. Jiang, W. Liu, L. Cao, and G. Long. Coupled collaborative filtering for context-aware recommendation. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 4172–4173, 2015.

[38] F. Li, G. Xu, and L. Cao. Coupled matrix factorization within non-iid context. In *Advances in Knowledge Discovery and Data Mining - 19th Pacific-Asia Conference, PAKDD 2015, Ho Chi Minh City, Vietnam, May 19-22, 2015, Proceedings, Part II*, pages 707–719, 2015.

[39] F. Li, G. Xu, L. Cao, X. Fan, and Z. Niu. Coupled group-based matrix factorization for recommender system. In *Web Information Systems Engineering - WISE 2013 - 14th International Conference, Nanjing, China, October 13-15, 2013, Proceedings, Part I*, pages 189–198, 2013.

[40] M. Li, J. Li, Y. Ou, Y. Zhang, D. Luo, M. Bahtia, and L. Cao. Coupled k-nearest centroid classification for non-iid data. *Trans. Computational Collective Intelligence*, 15:89–100, 2014.

[41] C. Liu and L. Cao. A coupled k-nearest neighbor algorithm for multi-label classification. In *Advances in Knowledge Discovery and Data Mining - 19th Pacific-Asia Conference, PAKDD 2015, Ho Chi Minh City, Vietnam, May 19-22, 2015, Proceedings, Part I*, pages 176–187, 2015.

[42] C. Liu, L. Cao, and P. S. Yu. Coupled fuzzy k-nearest neighbors classification of imbalanced non-iid categorical data. In *2014 International Joint Conference on Neural Networks, IJCNN 2014, Beijing, China, July 6-11, 2014*, pages 1122–1129, 2014.

[43] C. Liu, L. Cao, and P. S. Yu. A hybrid coupled k-nearest neighbor algorithm on imbalance data. In *2014 International Joint Conference on Neural Networks, IJCNN 2014, Beijing, China, July 6-11, 2014*, pages 2011–2018, 2014.

[44] X. Meng, L. Cao, and J. Shao. Semantic approximate keyword query based on keyword and query coupling relationship analysis. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014*, pages 529–538, 2014.

[45] G. Pang, L. Cao, and L. Chen. Outlier detection in complex categorical data by modeling the feature value couplings. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pages 1902–1908, 2016.

[46] G. Pang, L. Cao, L. Chen, D. Lian, and H. Liu. Sparse modeling-based sequential ensemble learning for effective outlier detection in high-dimensional numeric data. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th*

innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pages 3892–3899, 2018.

[47] G. Pang, L. Cao, L. Chen, and H. Liu. Unsupervised feature selection for outlier detection by modelling hierarchical value-feature couplings. In *IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain*, pages 410–419, 2016.

[48] G. Pang, L. Cao, L. Chen, and H. Liu. Learning homophily couplings from non-iid data for joint feature selection and noise-resilient outlier detection. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 2585–2591, 2017.

[49] G. Pang, H. Xu, L. Cao, and W. Zhao. Selective value coupling learning for detecting outliers in high-dimensional categorical data. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, pages 807–816, 2017.

[50] J. Shao, J. Yin, W. Liu, and L. Cao. Actionable combined high utility itemset mining. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 4206–4207, 2015.

[51] Z. She, C. Wang, and L. Cao. Cce: A coupled framework of clustering ensembles. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada.*, 2012.

[52] Y. Shi, W. Li, Y. Gao, L. Cao, and D. Shen. Beyond iid: Learning to combine non-iid metrics for vision tasks. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 1524–1531, 2017.

[53] Y. Song and L. Cao. Graph-based coupled behavior analysis: A case study on detecting collaborative manipulations in stock markets. In *The 2012 International Joint Conference on Neural Networks (IJCNN), Brisbane, Australia, June 10-15, 2012*, pages 1–8, 2012.

[54] Y. Song, L. Cao, X. Wu, G. Wei, W. Ye, and W. Ding. Coupled behavior analysis for capturing coupling relationships in group-based market manipulations. In *The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, Beijing, China, August 12-16, 2012*, pages 976–984, 2012.

[55] C. Wang and L. Cao. Modeling and analysis of social activity process. In L. Cao and P. Y. (eds.), editors, *Behavior Computing: Modeling, Analysis, Mining and Decision*, pages 21–35. Springer, 2012.

[56] C. Wang, L. Cao, M. Wang, J. Li, W. Wei, and Y. Ou. Coupled nominal similarity in unsupervised learning.

In *Proceedings of the 20th ACM Conference on Information and Knowledge Management, CIKM 2011, Glasgow, United Kingdom, October 24-28, 2011*, pages 973–978, 2011.

[57] C. Wang, X. Dong, F. Zhou, L. Cao, and C. Chi. Coupled attribute similarity learning on categorical data. *IEEE Trans. Neural Netw. Learning Syst.*, 26(4):781–797, 2015.

[58] C. Wang, Z. She, and L. Cao. Coupled attribute analysis on numerical data. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, pages 1736–1742, 2013.

[59] C. Wang, Z. She, and L. Cao. Coupled clustering ensemble: Incorporating coupling relationships both between base clusterings and objects. In *29th IEEE International Conference on Data Engineering, ICDE 2013, Brisbane, Australia, April 8-12, 2013*, pages 374–385, 2013.

[60] C. Wang, M. Wang, Z. She, and L. Cao. Cd: A coupled discretization algorithm. In *Advances in Knowledge Discovery and Data Mining - 16th Pacific-Asia Conference, PAKDD 2012, Kuala Lumpur, Malaysia, May 29 - June 1, 2012, Proceedings, Part II*, pages 407–418, 2012.

[61] S. Wang and L. Cao. Inferring implicit rules by learning explicit and hidden item dependency. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pages 1–12, 2017.

[62] Z. Xu, Y. Zhang, and L. Cao. Social image analysis from a non-iid perspective. *IEEE Trans. Multimedia*, 16(7):1986–1998, 2014.

[63] Y. Yu, C. Wang, Y. Gao, L. Cao, and X. Chen. A coupled clustering approach for items recommendation. In *Advances in Knowledge Discovery and Data Mining, 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part II*, pages 365–376, 2013.

[64] H. Zhang, Y. Zhao, L. Cao, and C. Zhang. Combined association rule mining. In *Advances in Knowledge Discovery and Data Mining, 12th Pacific-Asia Conference, PAKDD 2008, Osaka, Japan, May 20-23, 2008 Proceedings*, pages 1069–1074, 2008.

[65] Q. Zhang, L. Cao, C. Zhu, Z. Li, and J. Sun. Coupledcf: Learning explicit and implicit user-item couplings in recommendation for deep collaborative filtering. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 3662–3668, 2018.

[66] Y. Zhao, C. Zhang, and L. C. (Eds.), editors. *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*. Information Science Reference, 2009.

[67] Y. Zhao, H. Zhang, L. Cao, C. Zhang, and H. Bohlscheid. Combined pattern mining: From learned rules to actionable knowledge. In *AI 2008: Advances in Artificial Intelligence, 21st Australasian Joint Conference on Artificial Intelligence, Auckland, New Zealand, December 1-5, 2008. Proceedings*, pages 393–403, 2008.

[68] C. Zhu, L. Cao, Q. Liu, J. Yin, and V. Kumar. Heterogeneous metric learning of categorical data with hierarchical couplings. *IEEE Trans. Knowl. Data Eng.*, 30(7):1254–1267, 2018.

8.4 Previous tutorials: citation to an available example of work in the area

If it makes sense, information about previous tutorials given by the same presenters (title, conference, year number of attendees, etc.)

- Trong Dinh Thac Do and Longbing Cao. *Statistical Machine Learning of Large, Sparse and Multi-source Data*, PAKDD 2019.

- Longbing Cao. *Behavior analytics: methods and applications*, AAAI 2019.

- Longbing Cao, Philip Yu, Guansong Pang. *Behavior analytics: methods and applications*, KDD 2018.

- Liang Hu, Longbing Cao, Songlei Jian. *Non-IID Recommender Systems in Practice with Modern AI Techniques*, PAKDD 2018.

- Liang Hu, Longbing Cao, Jian Cao, Songlei Jian. *When Advanced Machine Learning Meets Intelligent Recommender Systems*, AAAI 2018.

- Longbing Cao, Philip Yu, Guansong Pang and Chengzhang Zhu. *Non-IID Learning*, KDD 2017.

- Longbing Cao, *Behavior Computing: Deep Behavior Analytics and Active Behavior Management*, PAKDD 2015.

- Longbing Cao. *Learning Non-IID Big Data*, CIKM 2014. The CIKM2014 tutorial addresses the same topic as in this tutorial proposal. It was the most popular tutorial at CIKM2014.

- Longbing Cao, *Non-IIDness Learning in Big Data*, PAKDD 2014. The PAKDD2014 tutorial addresses the same topic as in this tutorial proposal. It was the most popular tutorial at PAKDD2014.

- Longbing Cao, Philip S Yu, Can Wang, *Behavior Informatics: Modeling, Analysis and Mining of Complex Behaviors*, 4 Aug 2013, IJCAI 2013, Beijing, China. The room was full.

- Longbing Cao, Philip S Yu, Can Wang, *Behavior Computing: Complex Behavior Modeling, Analysis and Mining*, WI-IAT 2012, 4 Dec 2012, Macau, China.

- Longbing Cao, *Domain-Driven Data Mining: Empowering Actionable Knowledge Delivery*, PAKDD2009, over 30 audiences.

- Longbing Cao, *Agent-mining interaction and integration*, IEEE/WIC/ACM WI-IAT2007, USA, over 50 audiences.

8.5 Evidence of teaching experience

Speakers are very experience presenters, with numerous lectures and seminars delivered. The chief presenter Prof. Longbing Cao has delivered courses at both undergraduate and postgraduate levels in Australia, China, Italy, and France, and seminars and invited/keynote speeches at many conferences and countries. He has successfully delivered several tutorials at conferences including IJCAI, AAAI and KDD. Trong Dinh Thac Do has delivered seminars and presentations at several places and conferences, including IJCAI. He has also delivered some courses at undergraduate level in Vietnam and will deliver a tutorial in PAKDD 2019. Chengzhang Zhu has delivered seminars and presentations at several places and conferences, including IJCAI.

8.6 Evidence of scholarship in AI/computer science

In recent years, the tutorial lecturers have been awarded a series of scholarships in inventing new theories and application in non-IID learning, for instance, the following projects/grants awarded to Prof Cao that are relevant to this tutorial:

- Longbing Cao and Vipin Kumar. Deep analytics of non-occurring but important behaviours, DP190101079, 2019-2021.

- Longbing Cao, Philip S Yu, and Gerhard Weiss. Modelling and discovering complex interaction relations hidden in group behaviours in businesses, online and social communities, DP130102691, 2013-2015.

- Longbing Cao, Philip S Yu. Pattern Discovery of Discriminating Behaviour Associated with Hidden Communities, DP1096218, 2010-2012.

- Longbing Cao, Chengqi Zhang. Discovering Activity Patterns Driven by High Impacts in Heterogeneous and Imbalanced Data, DP0988016, 2009-2011.

- Longbing Cao. Domain-Driven Actionable Link Discovery, DP0773412, 2007-2009.

- Chengqi Zhang, Longbing Cao. Efficient Techniques for Mining Exceptional Patterns, DP0667060, 2006-2008.

- Chengqi Zhang, Longbing Cao, Ling Chen. Mining complex concurrency relationship patterns for dynamic customer/asset interaction modelling through novel industrial behaviour networks, LP120100566, 2012-2014, Partner Organization: Coates Hire Australia.

- Longbing Cao, Xingquan Zhu, Chengqi Zhang. Detecting significant changes in organisation customer interactions leading to non-compliance, LP100200774, 2010-2013, Partner Organization: Australian Federal Department of Human Services - Centrelink.

- Longbing Cao, Chengqi Zhang. Pattern Analysis and Risk Control of E-Commerce Transactions to Secure Online Payments, LP0989721, 2009-2011, Partner Organization: Westpac Banking Group.

- Chengqi Zhang, Longbing Cao, Dr MK Browne; Mrs YK Morrow; Mr R Schurmann; Mr PG Newbigin; Mr F Figueiredo; Mr Y Zhao (APDI). Data Mining of Activity Transactions to Strengthen Debt Prevention, 2007-2009, Partner Organization: Australian Commonwealth Government Agency Centrelink.

- Longbing Cao. Superannuation customer behavior analysis, AMP, 2010-2011.

- Longbing Cao. Detecting Incorrect Income Declaration, Centrelink, 2010.

- Longbing Cao. Health Insurance Behavior Analysis and Fraud Control, HCF and CMCRC, 4 full scholarships, 2008-2011.

- Longbing Cao (a CI in the project). Designing Out Crime Research Centre, Attorney General's Department (Crime Prevention Division), NSW Government, 2008-2010.