

CONTACT INFORMATION	<p>Department of Information Systems, Business Statistics, and Operations Management, School of Business and Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong</p> <p><a href="mailto:imyyang@ust.hk">imyyang@ust.hk</a></p>
EMPLOYMENT	<p><b>Hong Kong University of Science and Technology</b></p> <p><i>Assistant Professor,</i> July 2017-present School of Business and Management</p> <p><i>Affiliated Faculty,</i> July 2019-present Risk Management and Business Intelligence Program</p>
EDUCATION	<p><b>Northwestern University</b>, Evanston, Illinois Ph.D. in Computer Science Aug 2015</p> <p><b>University of Science and Technology of China</b>, Anhui, China B.S. in Computer Science July 2009</p>
RESEARCH INTERESTS	<p><b>Method Area:</b> Machine Learning, Natural Language Processing, Deep Learning <b>Applied Area:</b> Financial Technology, Marketing Analytics, Network Analytics</p>
JOURNAL PUBLICATIONS	<ol style="list-style-type: none"> <li>1. <b>Yi Yang</b> and Ramanath Subramanyam. Extracting actionable insights from text data: a stable topic model approach. <i>MIS Quarterly</i>, Forthcoming</li> <li>2. <b>Yi Yang</b>, Qin Yu, Yangyang Fan, and Zhongju Zhang. Unlocking the power of voice for financial risk prediction: A theory-driven deep learning design approach. <i>MIS Quarterly</i>, Forthcoming</li> <li>3. Allen H Huang, Hui Wang, and <b>Yi Yang</b>. Finbert: A large language model for extracting information from financial text. <i>Contemporary Accounting Research</i>, 2022</li> <li>4. <b>Yi Yang</b>, Kunpeng Zhang, and Yangyang Fan. sdtm: A supervised bayesian deep topic model for text analytics. <i>Information Systems Research</i>, 2022</li> <li>5. Zhuoyi Peng, <b>Yi Yang</b>, Liu Yang, and Kai Chen. Federated meta embedding concept stock recommendation. <i>IEEE Transactions on Big Data</i>, 2022</li> <li>6. Zhe Tang, <b>Yi Yang</b>, Wen Li, Defu Lian, and Lixin Duan. Deep cross-attention network for crowdfunding success prediction. <i>IEEE Transactions on Multimedia</i>, 2022</li> <li>7. <b>Yi Yang</b>, Kunpeng Zhang, and PK Kannan. Identifying market structure: A deep network representation learning of social engagement. <i>Journal of Marketing</i>, 86(4):37–56, 2022</li> <li>8. <b>Yi Yang</b>, Kunpeng Zhang, and Yangyang Fan. Analyzing firm reports for volatility prediction: A knowledge-driven text-embedding approach. <i>INFORMS Journal on Computing</i>, 34(1):522–540, 2022</li> <li>9. Fan Zhou, Kunpeng Zhang, Bangying Wu, <b>Yi Yang</b>, and Harry Jiannan Wang. Unifying online and offline preference for social link prediction. <i>INFORMS Journal on Computing</i>, 33(4):1400–1418, 2021</li> <li>10. <b>Yi Yang</b>, Shimei Pan, Jie Lu, Mercan Topkara, and Yangqiu Song. The stability and usability of statistical topic models. <i>ACM Transactions on Interactive Intelligent Systems (TiiS)</i>, 6(2):1–23, 2016</li> </ol>

11. Tiantian Zhu, Hongyu Gao, **Yi Yang**, Kai Bu, Yan Chen, Doug Downey, Kathy Lee, and Alok N Choudhary. Beating the artificial chaos: Fighting on spam using its own templates. *IEEE/ACM Transactions on Networking*, 24(6):3856–3869, 2016
  12. Fei Huang, Arun Ahuja, Doug Downey, **Yi Yang**, Yuhong Guo, and Alexander Yates. Learning representations for weakly supervised natural language processing tasks. *Computational Linguistics*, 40(1):85–120, 2014
- 
1. Yue Guo, **Yi Yang**, and Ahmed Abbasi. Auto-debias: Debiasing masked language models with automated biased prompts. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1012–1023, 2022
  2. Chengyu Chuang and **Yi Yang**. Buy tesla, sell ford: Assessing implicit stock market preference in pre-trained language models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 100–105, 2022
  3. John P Lalor, **Yi Yang**, Kendall Smith, Nicole Forsgren, and Ahmed Abbasi. Benchmarking intersectional biases in nlp. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3598–3609, 2022
  4. **Yi Yang**, Chewei Liu, and Sunil Mithas. Dual information technology strategy and firm performance: Insights from a deep learning approach. In *Academy of Management Proceedings*, volume 2021, page 12694. Academy of Management Briarcliff Manor, NY 10510, 2021
  5. Zhenyu He, Ce Li, Fan Zhou, and **Yi Yang**. Rumor detection on social media with event augmentations. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2020–2024, 2021
  6. Ahmed Abbasi, David Dobolyi, John P Lalor, Richard G Netemeyer, Kendall Smith, and **Yi Yang**. Constructing a psychometric testbed for fair natural language processing. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3748–3758, 2021
  7. Hanyu Duan, **Yi Yang**, and Kar Yan Tam. Learning numeracy: A simple yet effective number embedding approach using knowledge graph. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2597–2602, 2021
  8. Linyi Yang, Eoin Kenny, Tin Lok James Ng, **Yi Yang**, Barry Smyth, and Ruihai Dong. Generating plausible counterfactual explanations for deep transformers in financial text classification. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6150–6160, 2020
  9. Xinyi Wang and **Yi Yang**. Neural topic model with attention for supervised learning. In *International Conference on Artificial Intelligence and Statistics*, pages 1147–1156. PMLR, 2020
  10. Ting Zhong, Tianliang Wang, Fan Zhou, Goce Trajcevski, Kunpeng Zhang, and **Yi Yang**. Interpreting twitter user geolocation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 853–859, 2020
  11. Fan Zhou, Shengming Zhang, and **Yi Yang**. Interpretable operational risk classification with semi-supervised variational autoencoder. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 846–852, 2020
  12. Yu Qin and **Yi Yang**. What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 390–401, 2019

13. Fan Zhou, Bangying Wu, **Yi Yang**, Goce Trajcevski, Kunpeng Zhang, and Ting Zhong. Vec2link: Unifying heterogeneous data for social link prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1843–1846, 2018
14. Doug Downey, Chandra Bhagavatula, and **Yi Yang**. Efficient methods for inferring large sparse topic hierarchies. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 774–784, 2015
15. **Yi Yang**, Shimei Pan, Yangqiu Song, Jie Lu, Mercan Topkara, and JW Player. Improving topic model stability for effective document exploration. In *IJCAI*, pages 4223–4227, 2016
16. **Yi Yang**, Doug Downey, and Jordan Boyd-Graber. Efficient methods for incorporating knowledge into topic models. In *Proceeding of Empirical Methods in Natural Language Processing*, 2015
17. **Yi Yang**, Shimei Pan, Yangqiu Song, Jie Lu, and Mercan Topkara. User-directed non-disruptive topic model update for effective exploration of dynamic content. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 158–168, 2015
18. Kunpeng Zhang, **Yi Yang**, Aaron Sun, and Hengchang Liu. A systematic framework for sentiment identification by modeling user social effects. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, volume 2, pages 172–179. IEEE, 2014
19. Hongyu Gao, **Yi Yang**, Kai Bu, Yan Chen, Doug Downey, Kathy Lee, and Alok Choudhary. Spam ain’t as diverse as it seems: throttling osn spam with templates underneath. In *Proceedings of the 30th Annual Computer Security Applications Conference*, pages 76–85, 2014
20. **Yi Yang**, Alexander Yates, and Doug Downey. Overcoming the memory bottleneck in distributed training of latent variable models of text. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 579–584, 2013

## GRANTS

### Principal Investigator (PI)

1. HKUST-WeBank Joint Lab Project, HKD \$497,000, 2022-2023  
Title: Pre-training for Tabular Business Data
2. HKUST-Kaisa Joint Research Institute Seed Fund, HKD \$670,000, 2021-2022  
Title: Deep Learning in Financial NLP
3. Hong Kong Research Grants Council, Early Career Scheme, HKD \$458,000, 2019-2022  
Title: Beyond Text: The Information Role of Voice in Financial Market
4. University Initial Grant, HKD \$100,000, 2017-2020

### Co-Investigator (Co-I)

1. Hong Kong Research Grants Council, Theme-based Research Scheme, HKD \$20,659,000, 2018-2022  
Title: Contributing to the Development of Hong Kong into a Global Fintech Hub
2. National Natural Science Foundation of China, RMB \$2,570,000, 2020-2023  
Title: Technologies of Fusion Perception and Collaborative Processing for Behavior Big Data in Cyberspace
3. Hong Kong Research Grants Council, General Research Fund, HKD \$605,216, 2022-2023  
Title: Product Recommendations in Email Campaigns for Passive Customers

WORKING EXPERIENCE	<b>University of Illinois Urbana-Champaign</b> , Urbana, Illinois Postdoc Research Associate, College of Business	Aug 2015-June 2017
	<b>Amazon</b> , Seattle, Washington Software Engineer Intern, Amazon Web Services	Sep 2013-Dec 2013
	<b>IBM T.J. Watson Research</b> , Yorktown Heights, New York Research Intern, Intelligent Multimedia Interaction Division	May 2013-Sep 2013
TEACHING	<b>Hong Kong University of Science and Technology</b> , Instructor, ISOM3360: Data Mining for Business Analytics, (Undergraduate) ISOM3370: Big Data Technologies, (Undergraduate) ISOM5270: Big Data Analytics (MBA full-time and part-time) ISOM5270: Big Data Analytics (Master) ★ Finalist for <i>The Franklin Prize for Teaching Excellence</i> for Year 2021 ★ Finalist for <i>The Franklin Prize for Teaching Excellence</i> for Year 2018	July 2017-present
	<b>University of Illinois Urbana-Champaign</b> , Urbana, Illinois Instructor, BADM352: Database Design and Management ★ Awarded <i>Teachers Ranked as Excellent by Their Students</i> in Spring 2017 and Fall 2016.	Aug 2015-June 2017
STUDENTS SUPERVISION	<b>Postgraduate students:</b> - Yingpeng Zhu, PhD candidate, ISOM, Supervisor, 2017-2023 (expected) - Jiaxin Liu, PhD candidate, FinTech, Co-supervisor (with Prof. Kar Yan Tam at ISOM), 2020-present - Hanyu Duan, MPhil, ISOM, Co-supervisor (with Prof. Kar Yan Tam at ISOM), 2021-present - Yue Guo, PhD Candidate DSA, Co-supervisor (with Prof. Yangqiu Song at CSE) 2020-present - Zihan Meng, MPhil, CSE, Co-supervisor (with Prof. Qiong Luo at CSE) 2019-2021	
	<b>Undergraduate students:</b> - 30 undergraduate students that participated in HKUST Undergraduate Research Opportunities Program (UROP), July 2017- present	
SERVICE	<b>University Service</b> FT/PT MScISM Committee (Information Systems), 2022-2023 Search and Appointments Committee (Information Systems), 2021-2022 UG committee (Information Systems), 2020-2021 Technology Review Committee (TRC), 2022-present	
	<b>Editorial Board</b> Associate Editor, Information Technology and Management (2020-present)	
	<b>External Reviewer</b> Information Systems Research, Management Information Systems Quarterly, Management Science, INFORMS Journal on Computing, Decision Support Systems, Journal of the Association for Information Systems	