Resources

Reference papers for CS512 Spring 2018

Chapters on Text Mining (for the preparation of the 1st miterm exam)

Phrase Mining:

1. Marina Danilevsky, Chi Wang, Nihit Desai, Xiang Ren, Jingyi Guo, and Jiawei Han, "Automatic Construction and Ranking of Topical Keyphrases on Collections of Short Documents", Proc. of 2014 SIAM Int. Conf. on Data Mining (SDM’14), Philadelphia, PA, April 2014.
2. Ahmed El-Kishky, Yanglei Song, Chi Wang, Clare R. Voss, and Jiawei Han, "Scalable Topical Phrase Mining from Text Corpora", PVLDB 8(3): 305 - 316, 2015. Also, in Proc. 2015 Int. Conf. on Very Large Data Bases (VLDB’15), Kohala Coast, Hawaii, Sept. 2015. (covered@CS512)
3. Jialu Liu, Jingbo Shang, Chi Wang, Xiang Ren, Jiawei Han, "Mining Quality Phrases from Massive Text Corpora", in Proc. of 2015 ACM SIGMOD Int. Conf. on Management of Data (SIGMOD’15), Melbourne, Australia, May 2015 (covered@CS512)
4. Jingbo Shang, Jialu Liu, Meng Jiang, Xiang Ren, Clare R. Voss, Jiawei Han, “AutoPhrase: Automated Phrase Mining from Massive Text Corpora”, ArXiv: http://arxiv.org/abs/1702.04457 2017(covered@CS512) (will be published in IEEE TKDE 2018)

Entity Recognition and Typing

1. Xiang Ren, Ahmed El-Kishky, Chi Wang, Fangbo Tao, Clare R. Voss, Heng Ji, Jiawei Han, "ClusType: Effective Entity Recognition and Typing by Relation Phrase-Based Clustering", in Proc. of 2015 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’15), Sydney, Australia, Aug. 2015. (covered@CS512)
2. Xiang Ren, Zeqiu Wu, Wenqi He, Meng Qu, Clare Voss, Heng Ji, Tarek Abdelzaher and Jiawei Han, "CoType: Joint Extraction of Typed Entities and Relations with Knowledge Bases", in Proc. of 2017 World-Wide Web Conf. (WWW’17), Perth, Australia, Apr. 2017. (covered@CS512)
3. Xiang Ren, Wenqi He, Meng Qu, Lifu Huang, Heng Ji, and Jiawei Han, "AFET: Automatic Fine-Grained Entity Typing by Hierarchical Partial-Label Embedding", in Proc. of 2016 Conf. on Empirical Methods in Natural Language Processing (EMNLP’16), Austin, TX, Nov. 2016
4. Xiang Ren, Wenqi He, Meng Qu, Clare R. Voss, Heng Ji, Jiawei Han, "Label Noise Reduction in Entity Typing by Heterogeneous Partial-Label Embedding", in Proc. of 2016 ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD’16), San Francisco, CA, Aug. 2016
5. Zeqiu Wu, Xiang Ren, Frank F. Xu, Ji Li and Jiawei Han, "Indirect Supervision for Relation Extraction using Question-Answer Pairs", in Proc. of 2018 ACM Int. Conf. on Web Search and Data Mining (WSDM’18), Los Angeles, CA, Feb. 2018

Embedding and IT applications, Keyphrase inference

2. JianTang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. LINE: Large-scale information network embedding. in Proc. of 2015 Int. World-Wide Web Conf. (WWW’15), Florence, Italy, May 2015. (covered@CS512)
3. JianTang, Meng Qu, and Qiaozhu Mei. PTE: Predictive text embedding through large-scale heterogeneous text networks. in Proc. of 2015 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’15), Sydney, Australia, Aug. 2015
4. Meng Qu, Xiang Ren and Jiawei Han, "Automatic Synonym Discovery with Knowledge Bases", in Proc. of 2017 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’17), Halifax, Nova Scotia, Canada, Aug. 2017
5. Huan Gui, Qi Zhu, Liyuan Liu, Aston Zhang, Jiawei Han, "Expert Finding in Heterogeneous Bibliographic Networks with Locally-trained Embeddings", submitted for publication, 2018 (covered@CS512)
6. Jialu Liu, Xiang Ren, Jingbo Shang, Clare Voss and Jiawei Han, "Representing Documents via Latent Keyphrase Inference", in Proc. of 2016 Int. World-Wide Web Conf. (WWW’16), Montreal, Canada, April 2016 (covered@CS512)

Information Extraction by Pattern Discovery

1. Meng Jiang, Jingbo Shang, Taylor Cassidy, Xiang Ren, Lance Kaplan, Timothy Hanratty and Jiawei Han, "MetaPAD: Meta Pattern Discovery from Massive Text Corpora", in Proc. of 2017 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’17), Halifax, Nova Scotia, Canada, Aug. 2017. (covered@CS512)

Discovery of Multi-Faceted Taxonomy

1. Xin Rong, Zhe Chen, Qiaozhu Mei, Eytan Adar, "EgoSet: Exploiting Word Ego-networks and User-generated Ontology for Multifaceted Set Expansion", Proc. of 2016 ACM Int. Conf. on Web Search and Data Mining (WSDM’16), Feb. 2016 (covered@CS512)
2. Jiaming Shen, Zeqiu Wu, Dongming Lei, Jingbo Shang, Xiang Ren, Jiawei Han, "SetExpan: Corpus-based Set Expansion via Context Feature Selection and Rank Ensemble", in Proc. of 2017 European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD’17), Skopje, Macedonia, Sept. 2017 (covered@CS512)
3. Chao Zhang, Fangbo Tao, Xiusi Chen, Jiaming Shen, Meng Jiang, Brian Sadler, Michelle Vanni, and Jiawei Han, “TaxoGen: Constructing Topical Concept Taxonomy by Adaptive Term Embedding and Clustering”, submitted for publication, 2018 (covered@CS512)

Textcub and its construction

1. Fangbo Tao, Honglei Zhuang, Chi Wang Yu, Qi Wang, Taylor Cassidy, Lance Kaplan, Clare Voros, Jiawei Han, “Multi-Dimensional, Phrase-Based Summarization in Text Cubes”, Data Eng. Bulletin 39(3), Sept. 2016, pp. 74-84 (covered@CS512)
2. Fangbo Tao1, Chao Zhang, Xiushi Chen, Meng Jiang,Tim Hanratty, Lance Kaplan, Jiawei Han, "Doc2Cube: Automated Document Allocation to Text Cube via Dimension-Aware Joint Embedding", submitted for publication, 2018 (covered@CS512)

Chapters for the preparation of the 2nd midterm exam:

Chapters on Mining Heterogeneous Information Networks

1. Ting Chen and Yizhou Sun, "Task-Guided and Path-Augmented Heterogeneous Network Embedding for Author Identification," Proc. of 2017 ACM Int. Conf. on Web Search and Data Mining (WSDM’17), Cambridge UK, Feb. 2017 (covered@CS512)
3. Huan Gui, Jiulu Liu, Fangbo Tao, Meng Jiang, Brandon T. Norick, and Jiawei Han, “Large-Scale Embedding Learning in Heterogeneous Event Data”, in Proc. of 2016 Int. Conf. on Data Mining (ICDM’16), Barcelona, Spain, Dec. 2016.
4. Yizhou Sun, Yintao Yu, and Jiawei Han, "Ranking-Based Clustering of Heterogeneous Information Networks with Star Network Schema", Proc. 2009 ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD’09), Paris, France, June 2009 (covered@CS512)
5. Yizhou Sun, Jiawei Han, Peixiang Zhao, Zhijun Yin, Hong Cheng, Tianyi Wu, "RankClus: Integrating Clustering with Ranking for Heterogeneous Information Network Analysis", in Proc. 2009 Int. Conf. on Extending Data Base Technology (EDBT’09), Saint-Petersburg, Russia. Mar. 2009 (covered@CS512)
6. Yizhou Sun, Brandon Norick, Jiawei Han, Xifeng Yan, Philip S. Yu, and Xiao Yu, "Integrating Meta-Path Selection with User Guided Object Clustering in Heterogeneous Information Networks", Proc. of 2012 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’12), Beijing, China, Aug. 2012
7. Ming Ji, Jiawei Han, and Marina Danilevsky, “Ranking-Based Classification of Heterogeneous Information Networks”, Proc. of 2011 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’11), San Diego, Aug. 2011 (covered@CS512)
8. Ming Ji, Yizhou Sun, Marina Danilevsky, Jiawei Han, and Jing Gao, “Graph Regularized Transductive Classification on Heterogeneous Information Networks”, Proc. 2010 European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD’10), Barcelona, Spain, Sept. 2010.
9. Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, and Tianyi Wu, "PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks", Proc. of 2011 Int. Conf. on Very Large Data Bases (VLDB’11), Seattle, WA, Aug. 2011 (covered@CS512)
10. Yizhou Sun, Rick Barber, Manish Gupta, Charu Aggarwal and Jiawei Han, "Co-Author Relationship Prediction in Heterogeneous Bibliographic Networks", Proc. of 2011 Int. Conf. on Advances in Social Network Analysis and Mining (ASONAM’11), Kaohsiung, Taiwan, July 2011 [slides] (covered@CS512)
11. Yizhou Sun, Jiawei Han, Charu C. Aggarwal, and Nitesh Chawla, "When Will It Happen? Relationship Prediction in Heterogeneous Information Networks", Proc. 2012 ACM Int. Conf. on Web Search and Data Mining (WSDM’12), Seattle, WA, Feb. 2012.
12. Xiao Yu, Xiang Ren, Yizhou Sun, Quanquan Gu, Bradley Sturt, Urvashi Khandelwal, Brandon Norick, and Jiawei Han, "Personalized Entity Recommendation: A Heterogeneous Information Network Approach", in Proc. 2014 ACM Int. Conf. on Web Search and Data Mining (WSDM’14), New York, NY, Feb. 2014.
13. Xiang Ren, Jiulu Liu, Xiao Yu, Urvashi Khandelwal, Quanquan Gu, Lidan Wang, and Jiawei Han, "ClusaCite: Effective Citation Recommendation by Information Network-Based Clustering," in Proc. 2014 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’14), New York, NY, Aug. 2014 (covered@CS512)
14. Yizhou Sun, Jie Tang, Jiawei Han, Manish Gupta, and Bo Zhao, “Community Evolution Detection in Dynamic Heterogeneous Information Networks”, in Proc. 2010 KDD Workshop on Mining and Learning with Graphs (MLG’10), Washington D.C., July 2010 (covered@CS512)
15. Yizhou Sun and Jiawei Han, Mining Heterogeneous Information Networks: Principles and Methodologies, Morgan & Claypool Publishers, July 2012 (Series: Synthesis Lectures on Data Mining and Knowledge Discovery)
16. Chu Wang, Jiawei Han, Yuntao Jia, Jie Tang, Duo Zhang, Yintao Yu, and Jingyi Guo, "Mining Advisor-Advisee Relationships from Research Publication Networks", Proc. 2010 ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD’10), Washington D.C., July 2010. (covered@CS512)
17. Chen Chen, Xifeng Yan, Feida Zhu, Jiawei Han, and Philip S. Yu, “Graph OLAP: Towards Online Analytical Processing on Graphs”, Proc. 2008 Int. Conf. on Data Mining (ICDM’08), Pisa, Italy, Dec. 2008
18. Peixiang Zhao, Xiaolei Li, Dong Xin, and Jiawei Han, "Graph Cube: On Warehousing and OLAP Multidimensional Networks", Proc. of 2011 ACM SIGMOD Conf. on Management of Data (SIGMOD’11), Athens, Greece, June 2011 [slides]
19. Xiaoxin Yin, Jiawei Han, and Philip S. Yu, “Object Distinction: Distinguishing Objects with Identical Names by Link Analysis”, in Proc. 2007 Int. Conf. on Data Engineering (ICDE’07), Istanbul, Turkey, April 2007 (covered@CS512)
20. Xiaoxin Yin, Jiawei Han, Jiong Yang and Philip S. Yu, “Efficient Classification across Multiple Database Relations:A CrossMine Approach ”, IEEE Transactions on Knowledge and Data Engineering, 18(6): 770-783, 2006 (covered@CS512)

Chapter on Truth Finding
Chapter on Mining Spatiotemporal and Social Media Data

1. Zhijun Yin, Liangliang Cao, Quanquan Gu, and Jiawei Han, "A Probabilistic Model of Community-Based Latent Topic Analysis", ACM Transactions on Intelligent Systems and Technology (ACM TIST), 3(4), 2012, doi: 10.1145/2337542.2337548

2. Zhijun Yin, Liangliang Cao, Jiawei Han, Chengxiang Zhai, and Thomas Huang, "LPTRA: A Probabilistic Model for Latent Periodic Topic Analysis", in Proc. 2011 IEEE Int. Conf. on Data Mining (ICDM’11), Vancouver, Canada, Dec. 2011

3. Zhijun Yin, Liangliang Cao, Jiawei Han, Jiebo Luo, and Thomas Huang, "Diversified Trajectory Pattern Ranking in Geo-tagged Social Media", in Proc. of 2011 SIAM Conf. on Data Mining (SDM’11), Phoenix, AZ, Apr. 2011.

4. Zhijun Yin, Liangliang Cao, Jiawei Han, Chengxiang Zhai, and Thomas Huang, "Geographical Topic Discovery and Comparison", in Proc. of 2011 Int. World Wide Web Conf. (WWW’11), Hyderabad, India, Mar. 2011 (covered@CS512)

5. Chao Zhang, Yu Zheng, Xiuli Ma, Jiawei Han, "Assembler: Efficient Discovery of Spatial Coevolving Patterns in Massive Geosensory Data", in Proc. of 2015 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’15), Sydney, Australia, Aug. 2015

6. Chao Zhang, Jiawei Han, Lidan Shou, Jiajun Lu, and Thomas La Porta, "Splitter: Mining Fine Grained Sequential Patterns in Semantic Trajectories", PVLDB 7(9): 769-780, 2014 (Also, in Proc. 2014 Int. Conf. on Very Large Data Bases (VLDB’14), Hangzhou, China, Sept. 2014.)

7. Younghoon Kim, Jiawei Han, Cangzhou Yuan, "TOPTRAC: Topical Trajectory Pattern Mining", in Proc. of 2015 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’15), Sydney, Australia, Aug. 2015

8. Chao Zhang, Guangyu Zhou, Quan Yuan, Honglei Zhuang, Yu Zheng, Lance Kaplan, Shaowen Wang, Jiawei Han, "GeoBurst: Real-time Local Event Detection in Geo-tagged Tweet Stream", in Proc. of 2016 ACM SIGIR Conf. on Research & Development in Information Retrieval (SIGIR’16), Pisa, Italy, July 2016 (covered@CS512)

9. Jingjing Wang, Wenzhu Tong, Hongkun Yu, Min Li, Xiuli Ma, Haoyan Cai, Tim Hanratty, and Jiawei Han, "Mining Multi-Aspect Reflection of News Events in Twitter: Discovery, Linking and Presentation", in Proc. of 2015 IEEE Int. Conf. on Data Mining (ICDM’15), Atlantic City, NJ, Nov. 2015

10. Zhenhui Li, Jingjing Wang, and Jiawei Han, "Mining Event Periodicity from Incomplete Observations", in Proc. of 2012 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’12), Beijing, China, Aug. 2012 (slightly covered@CS512)

11. Zhenhui Li, Cindy Xide Lin, Bolin Ding, and Jiawei Han, "Mining Significant Time Intervals for Relationship Detection", Proc. of 2011 Int. Symp. on Spatial and Temporal Databases (SSTD’11), Minneapolis, MN, Aug. 2011.

12. Zhenhui Li, Bolin Ding, Jiawei Han, and Roland Kays, "Swarm: Mining Relaxed Temporal Moving Object Clusters", Proc. 2010 Int. Conf. on Very Large Data Bases (VLDB’10), Singapore, Sept. 2010.

13. Zhenhui Li, Bolin Ding, Jiawei Han, Roland Kays, and Peter Nye, "Mining Periodic Behaviors for Moving Objects", Proc. 2010 ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD’10), Washington D.C., July 2010

14. Xiaoyu Pan, Tantong W., Zhenhui Li and Jiawei Han, "Geo-Friends Recommendation in GPS-based Cyber-Physical Social Network", Proc. of 2011 Int. Conf. on Advances in Social Network Analysis and Mining (ASONAM’11), Kaohsiung, Taiwan, July 2011.

The following are not covered in 2018 class

Chapter on Stream Data Mining

1. Yixin Chen, Guozhu Dong, Jiawei Han, Benjamin W. Wah, and Jianyong Wang, "Multi-Dimensional Regression Analysis of Time-Series Data Streams", in Proc. 2002 Int. Conf. on Very Large Data Bases (VLDB’02), Hong Kong, China, Aug. 2002.

2. H. Wang, W. Fan, P. S. Yu, and J. Han, "Mining Concept-Drifting Data Streams using Ensemble Classifiers", in Proc. 2003 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’03), Washington, D.C., Aug. 2003.

3. C. Aggarwal, J. Han, J. Wang, and P. S. Yu, "A Framework for Clustering Evolving Data Streams", Proc. 2003 Int. Conf. on Very Large Data Bases (VLDB’03), Berlin, Germany, Sept. 2003.

4. Charu Aggarwal, Jiawei Han, Jianyong Wang, and Philip S. Yu, "On Efficient Algorithms for High Dimensional Projected Clustering of Data Streams", Data Mining and Knowledge Discovery, 10:251-272, 2005.

5. Jiawei Han, Yixin Chen, Guozhu Dong, Jian Pei, Benjamin W. Wah, Jianyong Wang, and Y. Dora Cai, "Stream Cube:An Architecture for Multi-Dimensional Analysis of Data Streams", Distributed and Parallel Databases, 18(2): 173-197, 2005.

6. Jing Gao, Wei Fan, and Jiawei Han, "A General Framework for Mining Concept-Drifting Data Streams with Skewed Distributions", in Proc. 2007 SIAM Int. Conf. on Data Mining (SDM’07), Minneapolis, MN, April 2007.

Related Knowledge on Networks
Chapter on Advanced Classification

1. Hong Cheng, Xifeng Yan, Jiawei Han, and Chih-Wei Hsu, "Discriminative Frequent Pattern Analysis for Effective Classification", in Proc. 2007 Int. Conf. on Data Engineering (ICDE’07), Istanbul, Turkey, April 2007.
2. Xifeng Yan, Hong Cheng, Jiawei Han, and Philip S. Yu, "Mining Significant Graph Patterns by Scalable Leap Search", Proc. 2008 ACM SIGMOD Int. Conf. on Management of Data (SIGMOD’08), Vancouver, BC, Canada, June 2008.
3. Hong Cheng, Xifeng Yan, Jiawei Han, and Philip S. Yu, "Direct Discriminative Pattern Mining for Effective Classification", in Proc. 2008 Int. Conf. on Data Engineering (ICDE’08), Cancun, Mexico, April 2008.
4. David Lo, Hong Cheng, Jiawei Han, SiauCheng Khoo, and Chengnian Sun, "Classification of Software Behaviors for Failure Detection: A Discriminative Pattern Mining Approach", in Proc. 2009 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’09), Paris, France, June 2009.
5. Jingbo Shang, Wenzhu Tong, Jian Peng, and Jiawei Han, "DPClass: An Effective but Concise Discriminative Patterns-Based Classification Framework", in Proc of 2016 SIAM Int. Conf. on Data Mining (SDM’16), Miami, FL, May 2016.

Chapter on Advanced Outlier Detection

1. Jonathan Kuck, Honglei Zhuang, Xifeng Yan, Hasan Cam, and Jiawei Han, "Query-Based Outlier Detection in Heterogeneous Information Networks", in Proc. of 2015 Int. Conf. on Extending Database Technology (EDBT’15), Brussels, Belgium, Mar. 2015.
2. Honglei Zhuang, Jing Zhang, George Brova, Jie Tang, Hasan Cam, Xifeng Yan and Jiawei Han, "Mining Query-Based Subnetwork Outliers in Heterogeneous Information Networks", in Proc. of 2014 IEEE Int. Conf. on Data Mining (ICDM’14), Shenzhen, China, Dec. 2014.
3. Manish Gupta, Jing Gao, Charu Aggarwal, and Jiawei Han, Outlier Detection for Temporal Data, Morgan & Claypool Publishers, 2014.
4. Manish Gupta, Jing Gao, and Jiawei Han, "Community Distribution Outlier Detection in Heterogeneous Information Networks", in Proc. of 2013 European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD’13), Prague, Czech, Sept. 2013, pp. 557-573.
5. Manish Gupta, Jing Gao, Yizhou Sun, and Jiawei Han, "Integrating Community Matching and Outlier Detection for Mining Evolutionary Community Outliers", in Proc of 2012 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’12), Beijing, China, Aug. 2012.

Chapter on Miscellaneous

1. Chi Wang, Jiawei Han, Qi Li, Xiang Li, Wen-Pin Lin, and Heng Ji, "Learning Hierarchical Relationships among Partially Ordered Objects with Heterogeneous Attributes and Links", Proc. 2012 SIAM Int. Conf. on Data Mining (SDM’12), Anaheim, CA, April 2012.
2. Chi Wang, Marina Danilevsky, Nihit Desai, Yinan Zhang, Phuong Nguyen, Thrivikrama Taula, and Jiawei Han, "A Phrase Mining Framework for Recursive Construction of a Topical Hierarchy", Proc. of 2013 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’13), Chicago, IL, Aug. 2013, pp. 437-445.
3. Chi Wang, Marina Danilevsky, Jialu Liu, Nihit Desai, Heng Ji, and Jiawei Han, "Constructing Topical Hierarchies in Heterogeneous Information Networks", Proc. 2013 IEEE Int. Conf.on Data Mining (ICDM’13), Dallas, TX, Dec. 2013.

Advanced Cluster Analysis

Please follow the textbook (3rd ed.) and the class slides

Outlier Analysis

Please follow the textbook (3rd ed.) and the class slides