



AALBORG UNIVERSITET

Big Data and AI at Daisy

Information for CSC Applicants

daisy

Research Leaders: Prof. Christian S. Jensen
Prof. Bin Yang

Center for Data-intensive Systems

Outline



- Denmark and Aalborg
- Aalborg University and Daisy
- Research focuses
- Projects
- Team
- Opportunities and Requirements

Denmark and Aalborg



- Denmark 
 - North European Country.
 - The happiest country.
 - Hans Christian Andersen' hometown.
- Aalborg
 - The capital of North Jutland.
 - The 3rd largest city in Denmark.
 - The happiest city in EU, according to a study from the European Commission in 2016.



Life in Aalborg



Oyster hunting 2019 summer



- Founded in 1974. Focus on Engineering and Technologies
- Engineering subjects rank No.1 in Europe and No.4 Worldwide (According to US News Ranking 2018 & 2019).
- No. 18 in the Top 100 of universities established within the past 50 years (According to Times Higher Education).
- No. 7 in the world within the field of Electronic and Electrical Engineering (According to Shanghai/ARWU World Rank).
- No. 37 in the top 50 of universities established within the past 50 years (According to QS).
- The best engineering program in Europe - a program that takes fourth place in the world! (According to the 2018 MIT report)

Daisy



- Center for Data-Intensive Systems
- Daisy ranks the 2nd best among all research groups in Europe according to publication performance in the top data management outlets in the recent 10-year period.
- Daisy Director, Prof. Christian S. Jensen,
 - ACM Fellow, IEEE Fellow,
 - Academician (院士) of the Academia Europaea, the Royal Danish Academy of Sciences and Letters, and the Danish Academy of Technical Sciences.
 - The highest h-index among all computer science professors in Denmark.

Institution	pubs	<i>frac</i>	inst/p	1 st	2 nd	aut
ETH Zurich	29	23	1.52	11	18	39
Aalborg University	26	14	1.96	11	15	18
Univ. of Edinburgh	25	15	2.16	5	20	16
INRIA Le Chesnay	23	12	2.35	20	3	33
Univ. of Athens	22	12	2.32	6	16	17
MPI Saarbrücken	17	12	1.71	2	15	9
CWI Amsterdam	14	12	1.43	5	9	15
Univ. of Munich	12	9	1.67	10	2	20

Affiliation analysis of database publications, SIGMOD Record 40(1):26-31 (2011)

Table 6: Top European research institutions

Data Analytics



- Spatio-temporal data analytics.
 - Machine learning and data mining on trajectories.
 - ◆ Mining significant semantic locations, co-movement patterns.
 - ◆ Driving behavior analysis, routing preference analysis.
 - Time-dependent and uncertain graphs.
 - ◆ High resolution modeling of travel costs, e.g., travel time fuel consumption.
 - ◆ Eco-routing, personalized routing.
 - Handling sparse data.
 - ◆ Spatial, temporal, and spatio-temporal sparseness.
 - Data-intensive routing.
 - ◆ Learning to route, learning to rank routes.

Data Analytics



- Time series analytics.
 - (Spatially-)Correlated time series analytics
 - ◆ Correlation analyses, forecasting.
 - Outlier detection
 - ◆ Unsupervised, automatic, and explainable.
- Graph data analytics.
 - Semi-supervised learning on graphs.
 - Graph convolution neural networks.
 - Graph embedding.

Data Management



- Spatio-temporal data management.
 - Indexing moving objects in Euclidean space, road networks, and indoor space.
 - ◆ TPR tree (1300 citations+), main memory indexing PGrid.
 - Query processing over moving objects.
 - ◆ Range queries, continuous queries, kNN.
 - ◆ Scalable platforms.
- Spatio-textual data management.
 - Indexing spatio-textual objects.
 - ◆ IR-Tree
 - Query processing over spatio-textual objects.
 - ◆ Rang queries, kNN, moving queries, region queries.
- Temporal data management.
 - Temporal analytics, e.g., different kinds of temporal aggregation
 - Temporal data modeling

Projects



- From national funds
 - DiCyPS: Data-Intensive Cyber-Physical Systems, DKK 67,140,000.
 - Obel Professor: free research on big data and AI, DKK 14,000,000.
 - Sapere Aude Research Leader: A Data-Intensive Paradigm for Dynamic, Uncertain Networks, DKK 5,882,000.
 - Analytics of time series in spatial networks, DKK 2,556,000.
 - Light-AI for Cognitive Power Electronics, DKK 2,996,278.
- From European Union funds
 - REDUCTION: utilize big traffic data to enable green transportation.
 - ◆ EUR 397,000
 - IT4BI-DC: Erasmus Mundus Joint Doctorate in Information Technologies for Business Intelligence – Doctoral College
 - ◆ Funding for ca. 100 Ph.D.s at 5 universities

Funding sources



- European commission.
- Independent research fund Denmark.
- The Danish agency for science and higher education.
- Innovation fund Denmark.
- The Obel family foundation.
- The Velux foundations.



Horizon 2020
European Union Funding
for Research & Innovation



INDEPENDENT
RESEARCH FUND
DENMARK



Ministry of Higher
Education and Science

Innovationsfonden



DET
OBELSKE
FAMILIEFOND

THE VELUX FOUNDATIONS

VILLUM FONDEN  VELUX FONDEN

Team



- Full professor
 - Christian S. Jensen and Bin Yang
- Associate professor
 - Chenjuan Guo
- Assistant professor (tenure-track)
 - Jilin Hu
- Postdoc
 - Tung Kieu: outlier detection, explainable AI.
- Phd students
 - Razvan-Gabriel Cirstea: time series analytics.
 - Simon Aagaard Pedersen: data intensive routing.
 - Ye Yuan: reinforcement learning for routing.
 - Sean B. Yang: machine learning on spatiotemporal data.
 - Tianyi Li: trajectory analytics.
- 21 master students



- CLAAUDIA
 - COMPUTE CLOUD

Type	Count	CPU type	Cores pr. node	Total cores	RAM pr. node	Total RAM	Additional specs.
Generic	11	AMD	64 cores	704	512 GB	~ 5,6 TB	InfiniBand
CLAAUDIA Managed	9	AMD	64 cores	576	512 GB	~ 4,6 TB	InfiniBand
Generic	19	AMD	64 cores	1216	1024 GB	~ 19,4 TB	
Generic	1	AMD	64 cores	64	2048 GB	~ 2,0 TB	
CLAAUDIA Managed	1	AMD	64 cores	64	2048 GB	~ 2,0 TB	
CLAAUDIA Managed	1	AMD	32 cores	32	256 GB	256 GB	
CLAAUDIA Managed	1	AMD	64 cores	64	512 GB	512 GB	
CLAAUDIA Managed	2	AMD	32 cores	64	512 GB	~ 1,0 TB	
CLAAUDIA Managed	21	AMD	16 cores	336	256 GB	~ 5,3 TB	
CLAAUDIA Managed	1	AMD	24 cores	24	512 GB	512 GB	
CLAAUDIA Managed	1	AMD	64 cores	64	512 GB	512 GB	
CLAAUDIA Managed	1	AMD	64 cores	64	256 GB	256 GB	
Generic	18	Intel	36 cores	648	384 GB	~ 6,7 TB	InfiniBand
Generic	10	Intel	36 cores	360	768 GB	~ 7,6 TB	
Search cloud	5	Intel	16 cores	80	128 GB	640 GB	16 TB SSD per node
Total	102			4360		~ 55 TB	

- AI CLOUD
 - ◆ Three NVIDIA DGX-2 (cost ca. 3 million DKK), each has 16 v100.
- COMMERCIAL CLOUD SERVICES
 - ◆ Infrastructure Cooperation (DeIC).
 - ◆ Azure cloud (Possible Access)

Working environment



Join Daisy!



- Phd, visiting phd students, and visiting scholar from CSC
 - Help applying for scholarships from China Scholarship Council.
 - Provide additional allowance when contributing to our ongoing projects.
 - Opportunities to visit top universities such as Oxford, MIT, Cornell, and NTU.
- Postdocs
 - Salary around 450,000 dkk per year before tax.
 - Possibility for promotion to assistant professor in the end of postdoc term, based on both research and teaching performance.
- Additional supports for participating international conferences and workshops.

Requirements



- Research Topics
 - Machine Learning
 - Data Mining
 - Data Management
 - Graph Neural Networks
 - Computer Vision
 - Algorithms
- Characteristics
 - Self-motivated, optimistic
 - Excellent academic record
 - Good knowledge of programming languages, e.g., Python, JAVA, or C++.
 - Proficiency in both oral and writing English, especially writing.

Contacts



- Bin Yang byang@cs.aau.dk
<http://people.cs.aau.dk/~byang/>
- Jilin Hu: hujilin@cs.aau.dk

Selected Publications



- Chenjuan Guo, Bin Yang, Jilin Hu, Christian S. Jensen, and Lu Chen. *Context-Aware, Preference-Based Vehicle Routing*. **The VLDB Journal**.
- Jilin Hu, Bin Yang, Chenjuan Guo, Christian S. Jensen, and Hui Xiong. *Stochastic Origin-Destination Matrix Forecasting Using Dual-Stage Graph Convolutional, Recurrent Neural Networks*. **ICDE 2020**.
- Simon Aagaard Pedersen, Bin Yang, and Christian S. Jensen. *Fast Stochastic Routing under Time-Varying Uncertainty*. **The VLDB Journal**.
- Tung Kieu, Bin Yang, Chenjuan Guo, and Christian S. Jensen. *Outlier Detection for Time Series with Recurrent Autoencoder Ensembles*. **IJCAI 2019**, 2725-2732.
- Jilin Hu, Chenjuan Guo, Bin Yang, and Christian S. Jensen. *Stochastic Weight Completion for Road Networks using Graph Convolutional Networks*. **ICDE 2019**, 1274-1285.
- Fan Xia, Bin Yang, Chengcheng Yu, Weining Qian, and Aoying Zhou. *Towards Longitudinal Analytics on Social Media Data*. **ICDE 2019**, 350-361.
- Chenjuan Guo, Bin Yang, Jilin Hu, and Christian S. Jensen. *Learning to Route with Sparse Trajectory Sets*. **ICDE 2018**, 1073-1084.
- Xiucheng Li, Kaiqi Zhao, Gao Cong, Christian S. Jensen, Wei Wei: Deep Representation Learning for Trajectory Similarity Computation. **ICDE 2018**, 617-628.
- Huiping Liu, Cheqing Jin, Bin Yang, and Aoying Zhou. *Finding Top-k Optimal Sequenced Routes*. **ICDE 2018**, 569-580.
- Jilin Hu, Bin Yang, Chenjuan Guo, and Christian S. Jensen. *Risk-Aware Path Selection with Time-Varying, Uncertain Travel Costs—A Time Series Approach*. **The VLDB Journal 27(2)**: 179-200 (2018).
- Bin Yang, Jian Dai, Chenjuan Guo, Christian S. Jensen, and Jilin Hu. *PACE: A PAtH-CENtric Paradigm For Stochastic Path Finding*. **The VLDB Journal 27(2)**: 153-178 (2018).

Selected Publications



- Huiping Liu, Cheqing Jin, Bin Yang, and Aoying Zhou. *Finding Top-k Shortest Paths with Diversity*. **TKDE 30(3)**, 488-502 (2018).
- Jian Dai, Bin Yang, Chenjuan Guo, Christian S. Jensen, and Jilin Hu. *Path Cost Distribution Estimation Using Trajectory Data*. **PVLDB 10(3)**: 85-96 (2016).
- Bin Yang, Chenjuan Guo, Yu Ma, and Christian S. Jensen. *Toward Personalized, Context-Aware Routing*. **The VLDB Journal 24(2)**:297-318 (2015).
- Jian Dai, Bin Yang, Chenjuan Guo, and Zhiming Ding. *Personalized Route Recommendation using Big Trajectory Data*. **ICDE 2015**, 543-554, Seoul, Korea, April 2015.
- Jinling Jiang, Hua Lu, Bin Yang, and Bin Cui. *Finding Top-k Local Users in Geo-Tagged Social Media Data*. **ICDE 2015**, 267-278, Seoul, Korea, April 2015.
- Bin Yang, Chenjuan Guo, Christian S. Jensen, Manohar Kaul, and Shuo Shang. *Stochastic Skyline Route Planning Under Time-Varying Uncertainty*. **ICDE 2014**, 136-147.
- Manohar Kaul, Raymond Chi-Wing Wong, Bin Yang, and Christian S. Jensen. *Finding Shortest Paths on Terrains by Killing Two Birds with One Stone*. **PVLDB 7(1)**:73-84. VLDB 2014.
- Bin Yang, Manohar Kaul, and Christian S. Jensen. *Using Incomplete Information for Complete Weight Annotation of Road Networks*. **TKDE 26(5)**:1267-1279.
- Bin Yang, Chenjuan Guo, and Christian S. Jensen. *Travel Cost Inference from Sparse, Spatio-Temporally Correlated Time Series Using Markov Models*. **PVLDB 6(9)**:769-780. VLDB 2013.