Please send CV to ccl-job@baidu.com   **Locations**: Bellevue WA or Beijing China   Baidu 百度

Research Intern – Cognitive Computing Lab & Seattle Research Institute

We invite applications for research interns throughout the years (Summer, Fall, and Spring), in Bellevue WA, Sunnyvale CA, and Beijing China. Research interns will focus on original basic research in theory, algorithms, statistics, deep learning, reinforcement learning, natural language processing, knowledge graph, security, or computer vision. Applicants must be currently pursuing PhD degrees in Computer Science or related fields (such as Statistics, Electrical Engineering, Mathematics, Operation Research etc.). Applicants must also demonstrate excellent potential in research, through (e.g.,) publications in premier journals or top CS conference proceedings. Typically, the duration of the internship will be 3 – 6 months.

Postdoctoral Researcher – Cognitive Computing Lab & Seattle Research Institute

We look for outstanding researchers with strong background in machine learning, statistics, applied mathematics, systems, databases, NLP, computer vision, security, theoretical computer science, etc. Our mission is to develop next generation cognitive computing technologies for better connecting billions of users to services. Our postdoctoral researchers are expected to focus on basic research in broad AI related fields. This would be an excellent opportunity for fresh PhD graduates in CS, Statistics, EE, Applied Math, etc., to spend 1 – 3 years in an industrial research environment to prepare for their long-term research careers either in academia or research labs.

Research Scientist – Cognitive Computing Lab & Seattle Research Institute

We look for outstanding research scientists with strong background in statistical machine learning, deep learning, systems, databases, NLP, knowledge, reasoning, and/or computer vision. Our mission is to develop next generation cognitive computing technologies for better connecting billions of users to services. As a research scientist at Baidu, you will be uniquely positioned in our team to work on very large scale industry problems and to push forward frontiers of cognitive computing technologies. Publications in premier conferences or journals are highly encouraged.

Qualifications:

- PhD in Computer Science, Statistics, Electrical Engineering, Mathematics, Operation Research, etc.
- Excellent publication record in major CS conferences and/or premier Stat/EE/SIAM journals. Examples are CVPR, FOCS, KDD, ACL, WWW, ICML, SIGMOD, JMLR, PAMI, IEEE Info. Theory, major statistics/mathematics journals, SIAM J. Computing, SIAM J. Optimization, Math. Program. etc.
- Strong analytical and problem-solving skills. Team player with good communication skills.

Recent Media Coverage (In Chinese):

1. http://science.china.com.cn/2020-03/16/content_41092885.htm
2. https://www.jiqizhixin.com/articles/2020-05-09-14
3. https://news.163.com/20/1224/12/FUK3A8IT00019OH3.html

Selected Recent Publications

1. **NeurIPS 2020**, Towards Better Generalization of Adaptive Gradient Methods
2. **NeurIPS 2020**, Optimal Prediction of the Number of Unseen Species with Reproducibility
3. **NeurIPS 2020**, Ratio Trace Formulation of Wasserstein Discriminant Analysis
4. **NeurIPS 2020**, Thunder: a Fast Coordinate Selection Solver for Sparse Learning
5. **NeurIPS 2019**, Re-randomized Densification for One Permutation Hashing and Bin-wise Consistent Weighted Sampling
6. **NeurIPS 2019**, Outlier Detection and Robust PCA Using a Convex Measure of Innovation
7. **NeurIPS 2019**, Towards Practical Alternating Least-Squares for CCA
8. **NeurIPS 2019**, Generalization Error Analysis of Quantized Compressive Learning
9. **NeurIPS 2019**, Möbius Transformation for Fast Inner Product Search on Graph
10. **NeurIPS 2019**, Random Projections with Asymmetric Quantization
11. **ICML 2020**, Optimal Estimator for Unlabeled Linear Regression
12. **ICLR 2021**, Learning Energy-Based Generative Models via Coarse-to-Fine Expanding and Sampling
13. **ICLR 2019**, On Random Deep Weight-Tied Autoencoders: Exact Asymptotic Analysis, Phase Transitions…
14. **IEEE Trans. Information Theory (2018)**, On the Trade-Off Between Bit Depth and Number of Samples for a …
15. **Journal of Machine Learning Research (2020)**, On Convergence of Distributed Approximate Newton Methods…
16. **Journal of Machine Learning Research (2020)**, Two-Stage Approach to Multivariate Linear Regression…
17. **Journal of Machine Learning Research (2018)**, A Tight Bound of Hard Thresholding
18. **Journal of Machine Learning Research (2018)**, Gradient Hard Thresholding Pursuit
19. **COLT 2020**, Bessel Smoothing and Multi-Distribution Property Estimation
20. **COLT 2020**, Nearly Non-Expansive Bounds for Mahalanobis Hard Thresholding
21. **AISTATS 2021**, Stability and Risk Bounds of Iterative Hard Thresholding
22. **AISTATS 2021**, Identification of Matrix Joint Block Diagonalization
23. **AISTATS 2021**, On the Faster Alternating Least-Squares for CCA
24. **AISTATS 2021**, One Sketch for All: Non-linear Random Features from Compressed Linear Measurements
25. **AISTATS 2021**, Principal Subspace Estimation Under Information Diffusion
26. **AISTATS 2020**, Adaptive Online Kernel Sampling for Vertex Classification
27. **AISTATS 2020**, An Inverse-free Truncated Rayleigh-Ritz Method for Sparse Generalized Eigenvalue…
28. **AISTATS 2020**, Solving the Robust Matrix Completion Problem via a System of Nonlinear Equations
29. **AISTATS 2019**, Graph to Graph: a Topology Aware Approach for Graph Structures Learning & Generation
30. **UAI 2020**, A Practical Riemannian Algorithm for Computing Dominant Generalized Eigenspace
31. **UAI 2019**, A Sparse Representation-Based Approach to Linear Regression with Partially Shuffled Labels
32. **AAAI 2021**, Unsupervised Cross-Domain Translation via Alternating MCMC Teaching
33. **AAAI 2021**, Fast and Compact Bilinear Pooling by Shifted Random Maclaurin
34. **AAAI 2021**, Energy-Based Probability Estimation with Variational Ancestral Langevin Sampler
35. **AAAI 2021**, A Blind Block Term Decomposition of High Order Tensors
36. **AAAI 2021**, Rejection Sampling for Weighted Jaccard Similarity Revisited
37. **AAAI 2020**, IVFS: Simple and Efficient Feature Selection for High Dimensional Topology Preservation
38. **AAAI 2020**, Distributed Primal-Dual Optimization for Online Multi-Task Learning
39. **AAAI 2020**, Meta-CoTGAN: A Meta Cooperative Training Paradigm for Improving Adversarial Text …
40. **AAAI 2019**, Sign-Full Random Projections
41. **AAAI 2019**, Multi-Agent Discussion Mechanism for Natural Language Generation
42. **NAACL 2021**, Cross-lingual Cross-modal Pretraining for Multimodal Retrieval
43. **ACL 2020**, Cross-Lingual Unsupervised Sentiment Classification with Multi-View Transfer Learning
44. **ACL 2020**, Learning Interpretable Relationships … Entities, Relations and Concepts via Bayesian …
45. **ACL 2019**, End-to-end Deep Reinforcement Learning Based Coreference Resolution
46. **NAACL 2019**, Integration of Knowledge Graph Embedding into Topic Modeling with Hierarchical ….
47. **EMNLP 2020**, A Predicate-Function-Argument Annotation of Natural Language for Open-Domain …
48. **EMNLP 2019**, On Efficient Retrieval of Top Similarity Vectors
49. **EMNLP 2019**, Reinforced Product Metadata Selection for Helpfulness Assessment of Customer Reviews
50. **EMNLP 2018**, Logician and Orator: Learning from the Duality between Language and Knowledge in …

51. **IJCAI 2019**, Coreference Aware Representation Learning for Neural Named Entity Recognition
52. **MLSys 2020**, Distributed Hierarchical GPU Parameter Server for Massive Scale Deep Learning Ads Systems
53. **SIGIR 2020**, Video Recommendation with Multi-Gate Mixture of Experts Soft Actor Critic
54. **KDD 2020**, Combo-Attention Network for Baidu Video Advertising
55. **KDD 2019**, Hierarchical Multi-Task Word Embedding Learning for Synonym Prediction
56. **KDD 2019**, MOBIUS: Towards the Next Generation of Query-Ad Matching in Baidu's Sponsored Search
57. **KDD 2019**, GCN-MF: Disease-Gene Association Identification by Graph Convolutional Networks…
58. **KDD 2018**, R2SDH: Robust Rotated Supervised Discrete Hashing
59. **WWW 2021**, Consistent Sampling Through Extremal Process
60. **WWW 2021**, MQuadE: a Unified Model for Knowledge Fact Embedding
61. **WWW 2021**, Cross-lingual Language Model Pretraining for Retrieval
62. **WWW 2020**, Extracting Knowledge from Web Text with Monte Carlo Tree Search
63. **WWW 2020**, Estimate the Implicit Likelihoods of GANs with Application to Anomaly Detection
64. **WWW 2020**, Improved Touch-screen Inputting Using Sequence-level Prediction Generation
65. **WWW 2020**, Kernel Sampling for Online Multi-Task Classification
66. **WWW 2019**, Product-Aware Helpfulness Prediction of Online Reviews
67. **WWW 2019**, Large Scale Semantic Indexing with Deep Level-wise Extreme Multi-Label Learning
68. **WSDM 2020**, Fast Item Ranking under Neural Network based Measures
69. **WSDM 2019**, Knowledge Graph Embedding Based Question Answering
70. **WSDM 2018**, Collaborative Filtering via Additive Ordinal Regression
71. **WSDM 2018**, Logician: A Unified End-to-End Neural Approach for Open-Domain Information Extraction
72. **SIGMOD 2021**, Agile and Accurate CTR Prediction Model Training for Massive-Scale Online Advertising System
73. **ICDE 2021**, TIRA in Baidu Image Advertising
74. **ICDE 2020**, SONG: Approximate Nearest Neighbor Search on GPU
75. **ICDE 2020**, Efficient Attribute-Constrained Co-Located Community Search
76. **CIKM 2020**, Sample Optimization for Display Advertising
77. **CIKM 2020**, A Reinforced Semi-Supervised Neural Network for Helpful Review Identification
78. **CIKM 2019**, AIBox: CTR Prediction Model Training on a Single Node
79. **CIKM 2019**, Large Margin Prototypical Network for Few-shot Relation Classification with Fine-grained…
80. **CIKM 2018**, Collaborative Multi-Objective Ranking
81. **CIKM 2018**, FastInput: Improving Input Efficiency on Mobile Devices
82. **SDM 2020**, An Advantage Actor-Critic Algorithm with Confidence Exploration for Open Information…
83. **FODS 2020**, Toward Communication Efficient Adaptive Gradient Method
84. **FODS 2020**, Classification Acceleration via Merging Decision Trees
85. **ICWSM 2021**, Textual Analysis and Timely Detection of Suspended Social Media Accounts
86. **CVPR 2021**, Patchwise Generative ConvNet: Training an Energy-Based Model from a Single Natural Image for…
87. **CVPR 2021**, Learning Deep Generative Models by MCMC Inference with Optimal Transport Correction
88. **CVPR 2021**, Generative PointNet: Deep Energy-Based Learning on Unordered Point Sets for 3D Generation,…
89. **ICCV 2019**, Temporal Structure Mining for Weakly Supervised Action Detection
90. **ECCV 2020**, Toward Faster and Simpler Matrix Normalization via Rank-1 Update
91. **ECCV 2020**, Learning Noise-Aware Encoder-Decoder from Noisy Labels by Alternating Back-Propagation
92. **Statistica Sinica (2020)**, Methods for Sparse and Low-Rank Recovery under Simplex Constraints
93. **Electronic Journal of Statistics (2020)**, Rate optimal Chernoff bound and application to community detection…
94. **Physical Review Letters (2020)**, Magnetic Noise Enabled Biocompass
95. **BIT Numerical Mathematics (2019)**, Perturbation analysis of an eigenvector-dependent nonlinear eigenvalue
96. **Linear Algebra and its Applications (2019)**, Perturbation analysis for matrix joint block diagonalization
97. **Numerical Linear Algebra with Applications (2019)**, Solving the general joint block diagonalization problem …
98. **Journal of Statistical Planning and Inference (2020)**, Projected tests for high-dimensional covariance matrices
99. **International Journal of Computer Vision (2020)**, Product Quantization Network for Fast Visual Search
100. **IEEE Trans. Pattern Analysis Machine Intelligence (2021)**, Unsupervised Cross-Domain Translation via…

Locations: Bellevue WA, Beijing or Shenzhen China. Please send CV to ccl-job@baidu.com