Cagri Ozcaglar

Staff Software Engineer, Machine Learning
Meta
Sunnyvale, CA

cagriozcaglar@gmail.com
LinkedIn
518-694-6920

Skills

- ***Research & Industry Experience***: Machine Learning Applications | Search and Recommendation Systems | Retrieval and Ranking | Generative AI applications | NLP | Deep Learning | RAG | User Intent Modeling | Notification Systems | Social Network Analysis.
- ***Languages***: Python | Java | C++ | Scala | PHP / Hack | Shell | Perl | MATLAB | R | Groovy.
- ***Modeling***: PyTorch | Tensorflow | Keras | Python (scikit-learn) | Amazon ML.
- ***Data Processing***: Spark | Hadoop | Apache Pig | AWS.
- ***Databases***: SQL | Presto SQL | HiveQL | MySQL | PostgreSQL | SparkSQL | Oracle 10g/11g.

Experience

- **Meta,** *Menlo Park, CA* 1/2022 - Present
  Staff Software Engineer, Machine Learning

  Team: Meta AI Search Engine 2024 - Present
  - Led the ranking component of Meta AI Search Engine with 10+ engineers. Built parts of retrieval and L1 / L2 ranking stack for Meta AI Search Engine.
  - Built and productionized QERS (Query Embedding from Result Set) embeddings to improve retrieval and ranking:
    - V1: Query QERS embedding is set to reciprocal-rank-weighted document embeddings. QERS embedding is used in one of dense retrieval calls, and led to 20% gain in Recall@1000. Follow-up iterations included i18n (US / IN / PK), all-country fallback logic, and using fresh GSERP data inputs.
    - V2: Query QERS embedding is predicted from Serpray embedding with a model, which increased query embedding coverage to 100%. This change led to 4% increase in recall@1000.
  - Launched L1 ranker with iterative improvements: GBDT model is served in L1, iterations include hyperparameter tuning, using multi-label approach (GSERP, Context-Relevance, Needs-Met labels), i18n, feature engineering. All L1 launches combined led to 93% lift in recall@100.
  - Led L2 ranker track and launched multiple versions with iterative improvements: First versions are DNN models with dense / embedding / sparse arch, using multi-label approach (GSERP, Context-Relevance, Needs-Met labels), i18n, feature engineering. Then moved to co-trained teacher-student model with the same labeling method, and then to MTML model with multiple tasks. All L2 launches combined led to 85% lift in recall@5.
  - Contributed to the migration of L2 ranker from Pyper (PyTorch for Personalization, for SWEs) to Native Pytorch (Accelerated PyTorch Framework (APF)), leading to 10x faster training, lower inference CPU utilization.
  - Built a prototype of D2D similarity based Query->Document expansion, used in sparse retrieval for link prediction, which led to 2% increase in recall@300, 1% decrease in failure rate in retrieval.
  - Mentored internal and new hires, supported them by finding ramp-up projects and long-term projects.

  Team: Growth Notifications (GN) AI 2022 - 2024
  - Tech lead for AI Foundations pod. Worked with EMs, PMs, XFNs for team-wide / org-wide goal alignment and planning.
  - Led a cross-org initiative to launch unconnected Personalized Reels notifications with GN AI team, Video team, engineers / PMs / managers from both teams, and integrity team. Contributed to end-to-end implementation (retrieval, generators, user-level and notif-level filtering and ranking), leading to DAP impact with neutral guardrails (turnoffs, want rate). This launch led to +600k DAP.
  - Led modeling codebase migration from Caffe to PyTorch for GN AI team, with topline parity, capacity and MAP gains for prod models.
  - Opened several venues such as date-wise partitioning, using shampoo optimizer, which led to model improvements with DAP wins.
  - Reduced Gmail Priority Email Spam rate in Email Notifications with a proxy unsubscribe+spam prediction model, in collaboration with Google Gmail team, GN Experience team.
  - Led a cross-functional team of 10 people from GN AI, GN Experience, UX, and PMs to drive SMS inventory expansion, adding new notification types E2E (UX, generators / retrieval, ranking), which led to DAP wins.
  - Designed and implemented Generative Notifications generated by Llama for comment story notification types.
  - Moved Logged-out-Push click model from GBDT to SparseNN model architecture, returning DAP and MAP wins.
  - Used Bayesian optimization (QuickBO) in notif-level ranking in email notifications, to balance growth metrics including DAP, turnoff, sent volume, leading to DAP / MAP wins and turnoff reduction.
  - Used pre-trained XLM-R text embeddings from MultiRay / TextRay for notification recipient, actor, notification content, and landing page content for Push notifications, leading to DAP increase and turnoff reduction.

- **LinkedIn**, *Sunnyvale, CA* 10/2016 - 11/2021

Senior Software Engineer, Machine Learning

Teams: Down-Funnel Optimization, Careers Relevance 2018 - 2021

- Team Tech lead as of 2021: Created and led a team charter for two-sided Recruiter & Jobs Search & Recommendation Marketplace optimization.
- Won Data All Star Team Award at LinkedIn for 2020.
- Led multiple efforts for improving key metrics and AI productivity in Job Recommendation system ranking:
  - Led and implemented Job Recommendation system ranking component migration from GLMix model to deep neural network models using Tensorflow. Contributed to all components including data preprocessing and preparation, modeling, serving and infra changes. This led to 2% lift in confirmed hires (key metric), and 14% lift in positive feedback ratio. Increased AI productivity (time it takes to build end-to-end model and productionize) by 7x (14 days→2 days).
  - Led a team of engineers for Tensorflow v1→v2 migration and Photon-Connect v1→v2 migration of Job Recommendation system ranking, and contributed to design and implementation. This led to another increase in AI productivity by 2x (2 days→1 day).
- Owned, led, designed, and implemented Unified Offline Ranking EvaluatiOn (OREO) framework for LTSC (LinkedIn Talent Solutions and Careers) org, and integrated it with LinkedIn's ProML ecosystem. Onboarded users from external teams / orgs to OREO.
- Led and implemented multiple efforts on all aspects of Job Search Relevance, including query recommendations, candidate selection, and ranking.
  - Led Guided Search project for making query recommendations for Search starters and Inline query suggestions in Job Search product. Designed Unified Guided Search flow for Guided Search. Implemented and productionized Seq2Seq-LSTM model with Bahdanau Attention for generating query suggestions with online inference on free-form queries in Search starter and Inline Query Suggestion use cases. Inline Query suggestions with title→skill recommendations resulted in 1.5% lift in click rate@k.
  - Improved Job Search candidate selection by adding title-to-title expansion using title similarity via title Word2Vec embeddings, high-quality skill-ID-based retrieval, and standardization-based skill-to-title expansion. These changes led to two successful MME models, which resulted in cumulative 1.4% lift in confirmed hires. Built a rule learning model to generate rule precedence, translated to Galene (based on Lucene).
  - Improved Job Search ranking with multiple changes, including: 1) Labeling strategy update (Premium / Basic job applies are equal-labeled), 2) Feature Engineering using Frame (Feature Management System): Added searcher-job skill-coverage percentage features to be used in ranking, 3) Using member / job skill / title topic embeddings in L2 ranking. Cumulatively, these changes resulted in 1.5% lift in confirmed hires.

Team: Talent Relevance 2016 - 2018

- Led a team of engineers and researchers in Contextual Search project for Recruiter product.
  - Designed the offline and online architecture for Top-N skill recommendations for Contextual Search, led software engineers, and launched the product.
  - Co-designed the system for Candidate Recommendations with Query Generation and launched the product.
  - Improved Recruiter Search ranking using Contextual skill match features and topic vector match features via LDA. Designed and implemented offline and online pipeline for generating these features. Launched ranking models with contextual skill match features, and filed a patent for topic vector match features.
- Improved Recruiter Search ranking using various ranking models.
  - Designed, implemented, and productionized Generalized Linear Mixed (GLMix) models for Recruiter Search ranking with Learning-to-Rank features, XGBoost model scores and tree interaction features, with a 4-people team of engineers. This model led to 8.5% / 5% / 2% lift in InMail Accept@1/@5/@25 respectively, compared to baseline XGBoost model.
  - Published and presented this work in WWW'19: Entity Personalized Talent Search Models with Tree Interaction Features.
- Contributed to the design and implementation of an E2E deep learning model training pipeline for Recruiter Search using Tensorflow, Keras, Spark, Scala.
- Contributed to Recruiter Search ranking model training flow improvements.
  - Designed and implemented an E2E Offline Feature Engineering Workflow which allows users to quantify the impact of adding new features to Recruiter Search Modeling pipeline with hypothesis tests on the key evaluation metric.
  - Contributed to the migration of Recruiter Search label generation pipeline from Hadoop / Pig stack to Spark / Scala stack.
- Published 3 articles (SIGIR'2018, CIKM'2018, WWW'2019), filed 10 patents, and published 1 blog post on AI Behind LinkedIn Recruiter Search and Recommendation Systems.

- **Amazon**, *Seattle, WA* 9/2013 - 9/2016
  Research Scientist

Team: Consumer Marketing Analytics

- Designed, built, and productionized ranking models for various business lines, product categories, channels, programs.
  - Product category purchase prediction models in 11 marketplaces using weighted logistic regression, which led to 20% lift in purchase rates and revenue overall.
  - Channel response prediction models (email / mobile) using weighted logistic regression, which are used in conjunction with other prediction models.
  - Mobile channel acquisition and engagement models, Amazon Mobile shopping app download prediction models for mobile acquisition. Live tests with mobile-adjusted product category purchase prediction models, compared to Mobile App First Sign-in

(FSI) rate of 0.3% for universal control set, targeting Mobile shopping app download prediction model based segments with BAU offer and $5 incentive offer, returned 413% and 1476% incremental lift on FSI rate, respectively.
    - Custom Audience Targeting: Targeting Amazon customers on Facebook, using product category purchase prediction models and deal seeker prediction models. In live tests, campaigns targeting product category segments drove 23.4% E%O, and campaigns targeting deal seeker segments along with product category segments drove 14.67% E%O, which are the largest ROI the social channels has seen at Amazon at the time.
    - Prime Free-Trial program sign-up prediction models using thresholded model ensembles. Using two different ranking methods in live experiments, customer segments selected based on Prime Free-Trial sign-up prediction models returned 94% and 74% higher Prime Free-Trial program sign-up rates compared to the baseline sign-up rate of 7%.
    - Various channel usage prediction models and customer life-cycle prediction models, including Kindle Cross-Platform reader download prediction models, Google channel reliance models, Amazon Attrition prediction models.
- Designed, built, and productionized causal prediction models for product lines and channels.
    - Designed, built, and productionized causal inference models to calculate the differential prediction of a customer to take an action after targeting. Various applications include direct mail targeting for streaming Prime Instant Video, making a purchase from Fashion product categories, signing up for Amazon Business program. Live tests for PIV streaming show that incremental models return a lift of 32% and 121% on the percentage of streamers respectively, compared to baseline model and random targeting.
    - Designed and built causal inference models with importance weighting, in order to measure the differential likelihood of a customer to take an action after targeting in case of treatment set selection bias. In live experiments, bias-corrected uplift models returned higher incremental response rate compared to treatment model and biased uplift model. Filed a patent.
- Onboarded new users to Predix, an automated predictive modeling platform.
- Organized Machine Learning Talk Series for Consumer Marketing organization as a monthly recurring event.
- Published articles, presented posters, gave talks in Amazon internal machine learning conferences.

- **Bank of America, Merrill Lynch**, *New York, NY* 8/2012 - 8/2013
  Software Developer

  Team: Equity Linked Technology
  - Designed and implemented software for processing TESS real time feeds within trading systems.
  - Designed and implemented a connector between Access Request Management (ARM) and RAM database.

Education

- Ph.D., Computer Science 2012
  Rensselaer Polytechnic Institute Troy, NY
- M.S., Computer Science 2008
  Rensselaer Polytechnic Institute Troy, NY
- B.S., Computer Science 2006
  Bilkent University Ankara, Turkey

Publications

- Full list available in Google Scholar. See a representative list below.
    - *C. Ozcaglar*, et al., *Contextual search ranking using entity topic representations*. Patent issued, 2022.
    - *C. Ozcaglar*, et al., *Entity Personalized Talent Search Models with Tree Interaction Features*. WebConf (WWW), 2019.
    - R. Ramanath, et al., *Towards Deep and Representation Learning for Talent Search at LinkedIn*. CIKM, 2018.
    - S. Geyik, et al., *Talent Search and Recommendation Systems at LinkedIn: Practical Challenges and Lessons Learned*. SIGIR, 2018.

Awards

- Data All Star Team Award at LinkedIn, 2021.
- Transaction Risk Management Systems Modeling Hackathon at Amazon, first place, 2015.
- Student Travel Award to attend IEEE BIBM 2010.
- Full scholarship awarded by Rensselaer Polytechnic Institute for graduate study, 2006 - 2012.
- Full scholarship awarded by Bilkent University for undergraduate education, 2002 - 2006.
- Top 0.01% in nationwide University Entrance Exam among 1.5 million candidates, 2002.
- Ranked 1st in the Mediterranean Region in two Turkish Math Olympiads, 2000, 2001.
- Won bronze medals in Turkish Secondary School Math Olympiads, 1998, 1999.