Avneesh Singh Saluja

Contact Information	620 Vincente Ave. Berkeley, CA 94707, USA	☎: +1-650-526-8792 ✉: asaluja@gmail.com : http://asaluja.github.io

Synopsis

Applied researcher in multimodal machine learning. My current focus is on how we can build systems that extract knowledge from and reason over long-form narrative video in order to benefit various applications in content understanding and generation. From leading research pods to shipping production systems, I enjoy owning the full lifecycle of turning ideas into impact.

Relevant Professional Experience

9/2018–present **Netflix**
Los Angeles & Los Gatos, CA
Staff Research Scientist & Area Lead for Media Understanding

Highlights include:
- **Media Foundation Model**: trained multimodal transformer (video, audio, text) on 100K+ titles with novel fusion and contextualization losses, yielding a model that powers 10+ applications.
- **Content Assistant**: designed, built, and deployed an internal chatbot (500+ MAUs) for gaining content insights from title-specific data (scripts, metadata, model predictions), accelerating content analysis across multiple promotional asset creation teams.
- **Synopsis Generation**: Built an LLM-based system (instruction tuning + RLHF) that ingests multi-modal source material to generate on-brand synopses, now serving 80% of member-facing descriptions.
- **Scene Boundaries**: novel scene boundary detection algorithm (patent pending) enabling automated title chunking and ad insertion for downstream processing.
- **SeqCLIP**: adapted CLIP-style model to narrative video by leveraging relevant internal data sources, with applications in video search and media metadata tagging.
- **Demand Prediction**: developed multiple representation learning models (graph-based, transfer learning) for content and entertainment-centric entities (shows, movies, talent, books, metadata), with primary applications in content valuation & programming.

10/2015–9/2018 **Airbnb**
San Francisco, CA
Senior Research Scientist, AI Lab

Founding member of the AI lab and senior tech lead for applied research projects across multiple company initiatives, with a focus on multimodal learning for user-generated content. Worked on various projects across search, relevance, growth, infrastructure, trust & safety, and customer service.

Before 2015 Research internships at eBay Inc., Microsoft Research, and IBM Research.

Education

Carnegie Mellon University, Pittsburgh, PA

Ph.D. in Electrical & Computer Engineering, September 2015

- Dissertation: "Low-dimensional Context-dependent Translation Models"
- Advisors: Chris Dyer & Ian Lane
- Affiliated with Language Technologies Institute, School of Computer Science
- Recipient, Bertucci Fellowship (College of Engineering), 2015
- Recipient, eBay Graduate Fellowship, 2014

Carnegie Mellon University, Pittsburgh, PA

M.S. in Electrical & Computer Engineering, May 2013

- Graduate coursework in statistics, machine learning, learning theory, speech recognition, probabilistic graphical models, structured prediction, optimization

Stanford University, Stanford, CA

B.S. *with Distinction* in Electrical Engineering, April 2007

- Signal processing concentration
- President's Award for Academic Excellence, Tau Beta Pi

Advisory Roles

11/2018–present **Altovita**, London, U.K.
AI Advisor
Providing continuous guidance around the company's core AI product

3/2017–5/2022 **Sentieo**, San Francisco, CA
NLP Advisor
Guiding the Data Science team on best practices and cutting edge approaches for their summarization, transcription, and search efforts

Publications

Journals

A. Saluja and Y. Zhang; Online Discriminative Learning for Machine Translation with Binary-valued Feedback. *Machine Translation*, Vol. 28 (2) pp. 69-90, 2014.

Conferences

S. Mehta, B. Azarnoush, B. Chen, **A. Saluja**, V. Misra, B. Bahani, and R. Kumar; Simplify-then-Translate: Automatic Preprocessing for Black-Box Translation. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, 2020.

C. Mitcheltree, V. Wharton, and **A. Saluja**; Using Aspect Extraction Approaches to Generate Review Summaries and User Profiles. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) - Industry Track*, 2018.

A. Saluja, C. Dyer, and S. Cohen; Latent Variable Synchronous CFGs for Hierarchical Translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

A.P. Parikh, **A. Saluja**, C. Dyer, and E.P. Xing; Language Modeling with Power Low Rank Ensembles. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.**(Best Paper Nominee)**

A. Saluja, H. Hassan, K. Toutanova, and C. Quirk; Graph-based Semi-Supervised Learning of Translation Models from Monolingual Data. In *Proceedings of the Association for Computational Linguistics (ACL)*, 2014.

A. Saluja, I. Lane, and Y. Zhang; Machine Translation with Binary Feedback: a Large-Margin Approach. In *Conference of the Association for Machine Translation in the Americas (AMTA)*, 2012.

A. Saluja, P. Sundararajan, and O.J. Mengshoel; Age-Layered Expectation Maximization for Parameter Learning in Bayesian Networks. In *Artificial Intelligence & Statistics (AIStats)*, 2012.

A. Saluja, I. Lane, and Y. Zhang; Context-aware Language Modeling for Conversational Speech Translation. In *Machine Translation Summit XIII*, 2011.

Workshops

G. Bhat, **A. Saluja**, M. Dye, and J. Florjanczyk; Hierarchical Encoders for Modeling and Interpreting Screenplays. In *3rd Workshop on Narrative Understanding*, 2021.

A. Saluja and J. Navrátil; Graph-based Unsupervised Learning of Word Similarities Using Heterogeneous Feature Types. In *TextGraphs 8: Graph-based Methods for Natural Language Processing*, 2013.

A. Saluja and B. Kveton; Semi-Supervised Learning with Cover Trees. In *Big Learning: 2011 NIPS Workshop on Parallel and Large-Scale Machine Learning*, 2011.

A. Saluja, F. Mokaya, M. Phielipp, and B. Kveton; Automatic Identity Inference for Smart Televisions. In *AAAI 2011 Workshop on Lifelong Learning*, 2011.

Preprints & Technical Reports

A. Saluja, C. Dyer, and J.D. Ruvini; Paraphrase-Supervised Models of Compositionality. *arXiv:1801.10293*. February 2018 (originally February 2015).

A. Saluja, M. Pakdaman, D. Piao, and A.P. Parikh; Infinite Mixed Membership Matrix Factorization. *arXiv:1401:3413*. January 2014.

Invited Talks

June 2016	OpenAir Conference, San Francisco: *Machine Learning in a Community-driven Marketplace*
October 2016	Facebook Inc., Menlo Park: *Neural Networks and NLP at Airbnb*
November 2016	ReWork Machine Intelligence Summit, New York: *Extracting Customer Insights at Airbnb*
November 2017	Open Data Science Conference, San Francisco: *Deep Learning and Language Processing at Airbnb*
April 2018	AI NextCon Silicon Valley, Santa Clara: *Deep Learning and Language Processing at Airbnb*
August 2019	1st AllenNLP Summit, Seattle: *AllenNLP at Netflix*
December 2020	NeurIPS Expo: *NLP at Netflix*

Professional Service

Co-organizer	Modern Machine Learning and Natural Language Processing Workshop held at NIPS 2014
Reviewer	ACL, NAACL, EACL, EMNLP, JMLR, IWSLT, AAAI

Teaching Experience	8/2014–12/2014	**18-751** Applied Stochastic Processes Teaching Assistant and Guest Lecturer
	1/2013–5/2013	**18-799M** Advanced Machine Learning Teaching Assistant and Guest Lecturer
Languages	Programming:	Python, SQL, Bash
	Human:	English, Hindi, Urdu, Nepalese, Spanish (fluent); Punjabi, Bengali (conversational); Arabic (basic)
Personal	Lived in 9 countries on 6 continents, with K-12 schooling in international schools U.S. Citizen MENSA International member Interests: trumpet, violin, jazz, soccer, cricket, and squash	
References	Available upon request	