

## Areas of Interest

### Big Data Analytics

Adobe Experience Cloud has over 50 petabytes of data that we collect on behalf of our customers. This data includes marketing and advertising events that we track based on page visits, emails clicked, and ad impressions, as well as customer profile data that we collect or ingest. Most of the data comes in real-time and needs to be processed and acted on in real-time as well.

Adobe Experience Cloud has to deal with the volume, variety and velocity challenges of big data and we are interested in turning research into distributed algorithms that work on large amounts of high dimensional sparse data sets. This will allow us to find insights such as the most important features that affect a certain outcome. This can also help us describe a set of users to solve problems like anomaly detection, marketing mix attribution, and detecting seasonal or trending patterns.

### Experimentation and Experience Design

Adobe Experience Cloud delivers trillions of experiences every year in the form of web pages, ad impressions, emails, and social posts. Therefore, a great area of interest for Adobe is understanding how to run large numbers of experiments implicative of the causal effects of multiple asynchronized stochastic – not random – treatments. Also of great interest are: better algorithms for search, recommendations and merchandizing that improve user experience, and directed experimentations that may involve the automated generation of layouts and content.

### Journey Optimization

Adobe Experience Cloud does not only deliver individual experiences. It is also used to optimize the customer journey across multiple channels and touchpoints. We are interested in how to predict or match the right message or offer to the right audience, at the right time, on the right channel. This involves problems like determining the right type, number and timing of push messages (email, SMS notifications, etc.) to send a person to optimize probability of that person reengaging with a brand.

### Customer Intelligence

In the Experience Cloud, we manage billions of consumer profiles on behalf of our customers, so it is important to understand who the customer is and what they want. Problems that fall in this area of interest include: algorithmically identifying users across devices and across channels based on tracked data about the users, predicting customers' propensities to take certain actions based on their past behavior, and predicting the lifetime value of a customer. Similarly, our research interests include grouping together similar users, by creating segments, or identifying look-alike users that are likely to engage in similar behavior. These users can then be targeted with the right experiences such as an offer or a recommendation.

### Content Intelligence

Adobe Experience Cloud powers websites, apps, emails, ads and delivers trillions of digital content every year. We are interested in research that focuses on understanding content, generation of content and content analytics with the goal of making it easy to generate, manage and deliver content in a highly-personalized, relevant and aesthetically-pleasing way. Different media types lend themselves to different kinds of problems (e.g, video ad duration, website layout, email subject line, etc.).

## Media Optimization

There are many common areas of interest between marketing and advertising, especially as it relates to consumer intelligence, content intelligence and experience delivery. The additional areas of research in advertising arise from the pricing of limited ad inventory and the cost of buying media and budget restrictions that need to be met in optimization. Since ads are purchased on third party sites, we are also interested in research that assesses the quality of media as it relates to fraud and band safety.

## Examples of Problems

### Big Data Analytics

- How do we use distributed algorithms that work on large amounts of high dimensional sparse data sets to find insights such as the most important features that affect a certain outcome or describe a set of users?
- What are good distributed techniques for anomaly detection and correlation detection? What are the contributing factors?
- What are good scalable methods for marketing mix attribution and contribution of marketing channels to observed outcomes?
- What are good techniques for detecting seasonal or trending patterns and leading indicators?
- *Data storytelling* — How can automatic analysis be blended with natural language generation (NLG) to produce data stories or narratives?

### Experimentation and Experience Design

- What is the most impression-efficient design of experiments?
- What are good methodologies to inference the causal effects of multiple asynchronized stochastic (not random) treatments (e.g., media touches)?
- How can we recommend good content or creatives for testing, based solely on characteristics of the content or creative and on the users who will see the creative?
- How can we automatically suggest and assemble good creatives from a repository of assets?
- How do we recommend products to maximize conversion or engagement given product catalogs, video catalogs, websites, click streams, and conversion data?
- What is the best way to conduct automatic A/B testing of digital experiences (i.e., automatically generating variations of layout or content) while adhering to aesthetics and brand requirements?
- How do we best analyze multiple concurrent A/B tests (not necessarily of the same duration) for the same user experience or journey in order to avoid confounding effects?
- How do we incorporate, into product search algorithms, multiple requirements of:
  - a) relevance to the end user need
  - b) the marketer's need to optimize for key revenue and profitability metrics
  - c) trends based on seasonality, weather and buzz about specific product categories/brands/products
  - d) the device on which the search is being performed and the location/immediacy that it connotes
- What are the best techniques for engaging with end users when they do not want to download apps on their devices or are reluctant to share their personal information?

## Journey Optimization

- How do we predict or learn to match the right message or offer (or next-best offer) to the right audience, at the right time, on the right channel?
- How can you determine the right type, number and timing of push messages (email, SMS notifications, etc.) to send a person to optimize probability of that person reengaging with a brand?
- How can we help the author of an email/notification improve the open rate by suggesting words for the subject line?
- How can we help the author of an email/notification find the right audience by analyzing the content of the email and past recipient behaviors?
- How do we predict how many emails/notifications people are willing to receive over a certain period of time without running the risk they will unsubscribe?
- How can we increase the chance a recipient will read an email/notification by sending it at the right time of day/day of week?
- *Location detection* — How can we predict the current and future location of a customer from mobile analytics data?

## Customer Intelligence

- How can we algorithmically identify users across devices and across channels based on tracked data?
- How can we predict and evaluate the overall lifetime value of a customer?
- How do we conduct probabilistic stitching of cross-device identities (e.g, identify that a person is the same across desktop and mobile device)?
- How can we automatically segment users based on cross-channel behavioral data?
- How can we estimate the lifetime value of a user?
- How can we automatically create segments and look-alike scores for users?
- *Audience intelligence* — How can we predict the behavior of a user or an audience segment?
- *Audience planning* — How can we recommend the right audience for a specific offer or a product promotion?

## Content Intelligence

### Content Understanding

- How can we automatically tag images and connect relevant images to written content?
- How can we construct and leverage ontologies (i.e., taking advantage of knowledge regarding concept hierarchies) in the tagging of assets for tag enrichment and visualization?
- How do we automatically convert paper or digital forms into responsive forms for mobile devices? How can we also include recognition of data entry fields and display tags?
- How can we predict missing content on a page? Automatically author content? Tailor relevant content to different screens?

### Content Generation

- How can we automatically create great layouts for websites based on the content (text, video, images, navigation, cart, etc.) that is needed on the site?
- How can we evolve or optimize website layout and content based on how users engage with the site (e.g, adding a link to content that is often searched for or moving a highly watched video to the top)?

- How can we be more intelligent (beyond responsive) about automatically optimizing content, layout, and site structure for different screen sizes (from smartphones to big screens)?
- How can we help the author of an experience by suggesting content (text, images, video) which complements the content that is already part of the experience (e.g., by identifying mood or topic from text and images and matching new content with existing content)?

## **Content Analytics**

- How can we determine quality of creative assets and quantified/predictive impact?
- How can we make automatic suggestions to text for improved Google rankings?
- How do we determine a customer's psychographics, and how do we find content that matches those psychographics?
- How do we find content relevant to a community of users? How can we suggest topics for new stories based on available content and external signals (like social media) and suggest available material that can be repurposed for this?
- How do we drive and measure engagement with video content? What are the factors that affect engagement?
- What are good techniques for discovering or recommending video content and video channels?
- How can publishers optimize their content catalog?
- What is the right number and duration of ads for video content? How are ads affected by the surrounding content? How should advertisers value the context around the video advertising?
- How do we analyze the use of augmented reality (AR) and virtual reality (VR) experiences to measure engagement? How would 360-degree VR video best be used to measure engagement and customer satisfaction?

## **Media Optimization**

- How can we do budget allocation, coordination, and timing of advertisements in a multichannel setting?
- What are good marketing models that combine offline and online consumer behavior?
- How do we build scalable, predictive models and optimization algorithms for targeting, bidding, and pacing in a real-time bidding marketplace?
- How do we predict consumers' buying propensity using cross-device data aggregation and analysis?
- What is the effect of price and inventory information on online ads based on consumer buying decisions?
- What are the best methods for attributing revenue to multiple marketing touchpoints and the frameworks to evaluate these methods?
- What are good techniques for revenue and cost estimation for long-tail biddable entities in online advertisements?
- How do we measure ad creative effectiveness in different channels and across devices?
- How can we automatically generate ad copy from data (e.g., creating an ad using a hotel description or product article)?
- What are good techniques to measure the effectiveness of online video advertising? How should advertisers spread their buys between branding and performance? Between television advertising and online video advertising?
- How do you detect fraudulent ad inventory and brand safety (i.e. delivery of ads in the right context)? How can we detect fraud in video tagging and video ad consumption?

## Sample Funded Proposals

### *Big Data Analytics*

- Sublinear Algorithms for Contextual Anomaly Detection and Attribution  
S. Muthukrishnan, Rutgers University
- Causal Inference Meets Machine Learning  
Ramesh Johari, Stanford University
- Experimentation and Attribution via Information-Directed Sampling  
Benjamin Van Roy, Stanford University
- Instance-Specific Explanations for Machine Learning Predictions  
Sameer Singh, UC Irvine
- Deep Inverse Reinforcement Learning for Multimodal Data Storytelling  
William Wang, University of California, Santa Barbara

### *Experimentation and Experience Design*

- Machine Learning for Causal Inference: Experiments, Targeting, and Evaluation  
Jasjeet Sekhon, UC Berkeley
- Off-Policy Policy Evaluation for Marketing  
Philip Thomas, University of Massachusetts Amherst
- Graph Trend Filtering for Recommendation  
Alexander Smola, Ryan Tibshirani, Carnegie Mellon University
- Improving the Transparency of Sequential Recommenders  
Ben Shneiderman, Catherine Plaisant, University of Maryland
- Spatial-Temporal Anomalous Pattern Detection (and A/B Testing)  
Edward McFowland III, University of Minnesota

### *Journey Optimization*

- Visual Analytics for Event Sequence Recommendation: EventAction  
Ben Shneiderman, Catherine Plaisant, University of Maryland
- The Impact of Latent Keyword Structure on Customers' Path to Purchase  
Hongshuang (Alice) Li, Indiana University
- Effects of Mobile Push Notifications on Purchase Behavior: The Role of Message Frequency, Timing and Content Framing  
Randolph E. Bucklin, UCLA (Anderson)
- Sequential Recommendations  
Ramesh Johari, Stanford University
- Strategic Timing of Emails  
Tauhid Zaman, MIT

### *Customer Intelligence*

- Cross-Device Graph-Based Entity Resolution  
Lise Getoor, UC Santa Cruz
- Algorithms for Exploratory Clustering at Massive Scale  
Dimitris Achlioptas, University of California, Santa Cruz

- Cohort Selection in Advertising  
Ashish Goel, Kamesh Munagala, Stanford University, Duke University
- Learning Multi-resolution Representations of Customer Behavior from Clickstream Data  
Zachary C. Lipton, Carnegie Mellon University
- Learning Individual Predictive Choice Models  
Padhraic Smyth, UC Irvine

## *Content Intelligence*

- Scalable, Adaptive Media Summarization  
Stefanie Jegelka, MIT
- Photo Feature X Is Worth Y Dollars: Investigating The Economic Impact Of Images in E-Commerce Via Deep Learning  
Dokyun Lee, Param Vir Singh, Shunyuan Zhang, Carnegie Mellon University
- KPI-Driven Content Understanding and Generation With Applications to Digital Marketing  
Zhi Wei, New Jersey Institute of Technology
- Videolyzer: An Interactive Platform for Making Persuasive Videos  
Wai-Tat Fu, Brittany Duff, Sanorita Dey, University of Illinois, Urbana Champaign
- Deep Learning for Understanding Documents  
Jamie Murdoch, Bin Yu, UC Berkeley

## *Media Optimization*

- Large Scale Experiments on the Efficacy of Digital Advertising  
Kinshuk Jerath, Miklos Sarvary & Ryan Dew, Columbia University
- Combining Machine Learning with Randomized Field Experiments to Improve Mobile Advertising  
Anindya Ghose, BeiBei Lei, NYU and Carnegie Mellon University
- Real-time Detection of Online Click & Display Advertising Exchange Fraud  
Lwman Akoglu, Tudor Dumitras, Carnegie Mellon University
- Advertising Portfolio Optimization Under Uncertainty  
Georgia Perakis, MIT Sloan
- Designing Effective Ad-Video Delivery Strategies  
Soumya Sen, Gordon Burch, Carlson School of Management - University of Minnesota