A Data-Driven Approach to Predict an Individual Customer's Call Arrival in Multichannel Customer Support Centers

Somayeh Moazeni

Stevens Institute of Technology

Hoboken, NJ, USA

smoazeni@stevens.edu

Joint work with **Rodrigo Andrade**

2018 IEEE International Congress on Big Data

San Francisco, CA, USA

July 4, 2018

Contact Center Multi-Channel Data

- Contact centers provide firms with the opportunity to collect rich customer interaction data from multiple channels.
- Analyzing these big datasets and developing accurate predictive models for customer behavior are essential to design and optimize business processes.
- Call forecasting: one of the three fundamental challenges in the management of call centers.

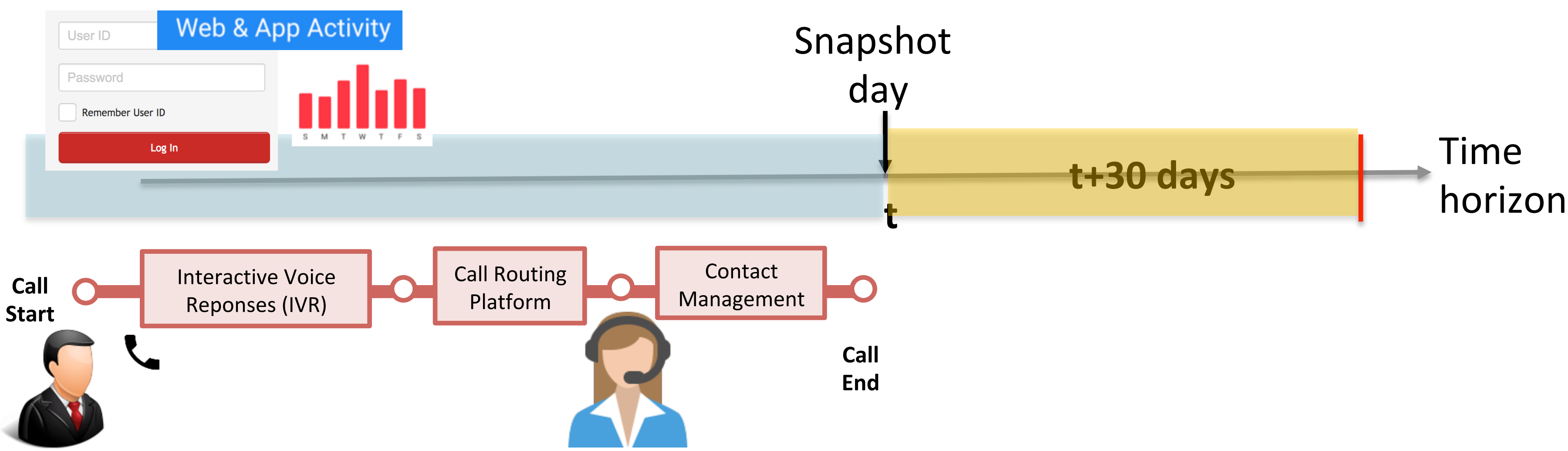
THE CHALLENGE:

Learning patterns in policyholders' interactions with a contact center and predicting future behavior of a specific customer.

Contact Center Multi-Channel Data

Leveraging **multichannel data** to **predict future telephone queries** by an **individual customer** and to examine the effect of **past Web-based contacts** by a customer on his future calls.

- Goal: to develop a feature-based model to predict the likelihood that a customer will call within the next thirty days.



Outline

- Related Literature
- Contact Center Data
- Data analysis
  - Feature Engineering
  - Data Exploration
- Predictive Model
- Scalable Data Analytics Method
- Results
  - Feature Selection
  - Model Performance
- Concluding Remarks

Related Literature

Target Variable: Call Arrival Counts

Univariate Models

References	Data
Brown et al. (2005)	Call Center of Israel's Bank.
Avramidis et al. (2004)	Bell Canada Call Center
Weinberg et al. (2007)	North American commercial bank.
Shen and Huang (2008a)	Northeastern U.S. bank call center.
Shen and Huang (2008b)	Northeastern U.S. financial firm.
Taylor (2008)	Retail bank in the UK
Taylor (2012)	NHS in England and Wales, and UK credit card company
Ibrahim and L'Ecuyer (2013)	Canadian Telecom. company

Models used: variablities of Doubly Stochastic Poisson Processes, and autoregressive models such as ARIMA

Multivariate Models

References	Soyer and Tarimcilar (2008)	Aldor-Noiman et al. (2009)
Data	Consumer electronics producer	Israel Telecom company
Model(s)	Two-way multiplicative Bayesian model.	Mixed Poisson Process and GLMM
Features	Media dollars, print media type (weekly or monthly), offer type	6 for days of the week, 8 for billing cycle

Target Variable: Call Outcome

Multivariate Models

Outbound Call Center	Inbound Call Center
Moro et. al. (2014)	Our Study
Data: Portuguese retail bank	Data: U.S. Insurance Comp.
Models: • Logistic regression • Decision Trees • Neural Network • SVM	Models: • Mixed-effect logistic model
N. of features: 150 Commonly used bank client and product attributes, and generic social and economic indicators	N. of features: 129 Features related to the call record, business, policy, product, and to customer profile and behavior. Additionally considers the weather as exogenous variable.

Related Literature

- The literature on call center predictions primarily focused on estimating the **intensity of call arrivals** to the call center based on **historical telephone queries.**
- Our paper focuses on:
  - customer-level predictions, and
  - includes features characterizing the customer’s past contacts via both Web and telephone channels
  - uses a rich set of features: contact reasons
  - relies on the Lasso method

Contact Center Data

- Data recorded from a major U.S. insurance firm
- January 1, 2015 to December 31, 2015.
- The data includes **35, 806, 207** transactions between **7, 463, 600** policyholders and the insurance firm.
- Transaction-level dataset consists of 7 attributes:

Attribute	Description
Event ID	Unique # to identify transactions corresponding to a contact event
Event Time-stamp	Date and time of contact
Contact Channel	The medium used for the contact
Contact Reason Type	High-level reason for the contact
Contact Reason Subtype	Detailed reason for the contact
Participant Type	Role of the participant who contacts
Policy ID Attributes	Attributes to identify a policyholder

- **Web** Transactions: Firm Website Account (69.2%)
- **Telephone** Transactions: Customer (45.4%), Agent (18.2%)

Contact Center Data

Use of Multiple Venues:

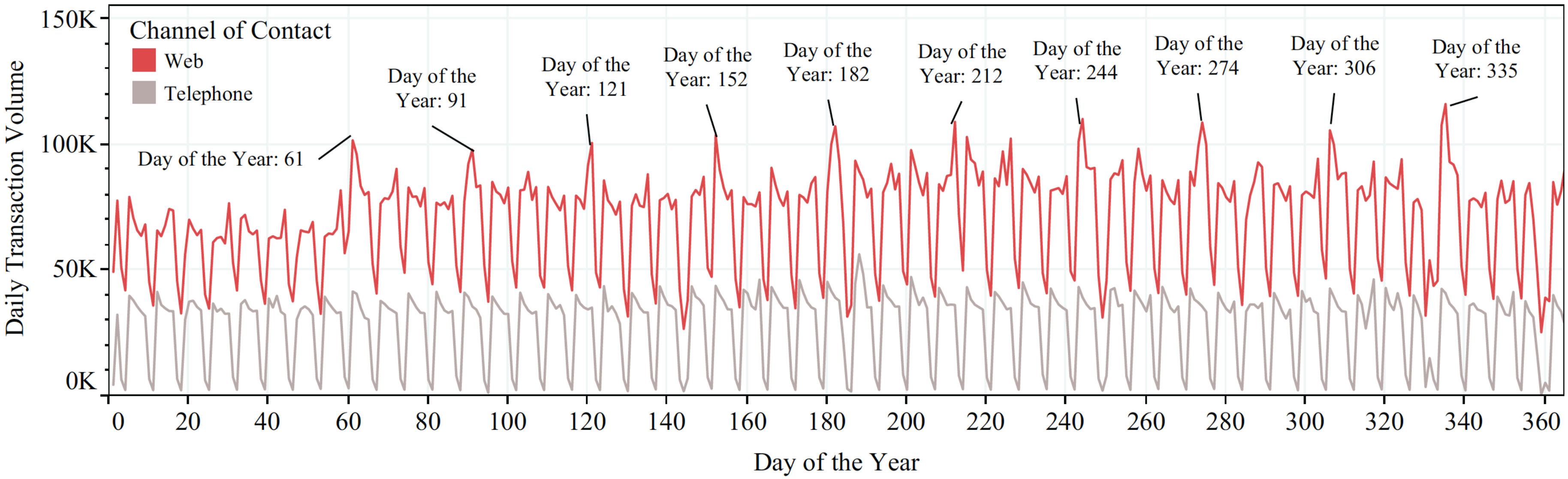
- Number of **Web** transactions = **25,833,965 (72.15%)**
- Number of **Telephone** transactions = **9,972,252 (27.85%)**

Usage	Policy Number Count	%	Transactions Count	%
Only **Web**	3,552,632	47.6%	18,518,930	51.7%
Only **Telephone**	2,274,760	30.5%	5,302,751	14.8%
Telephone and **Web**	1,636,208	21.9%	11,984,536	33.5%
Total	**7,463,600**		**35,806,217**	

- Larger number of policyholders have changed their medium of contact from **Telephone** to **Web** than from **Web** to **Telephone**:
  - **Web** to **Telephone**: 733, 751 policyholders
  - **Telephone** to **Web**: 1,466,620 policyholders

Contact Center Data

Daily volume of transactions per (Web, Telephone) during 2015:



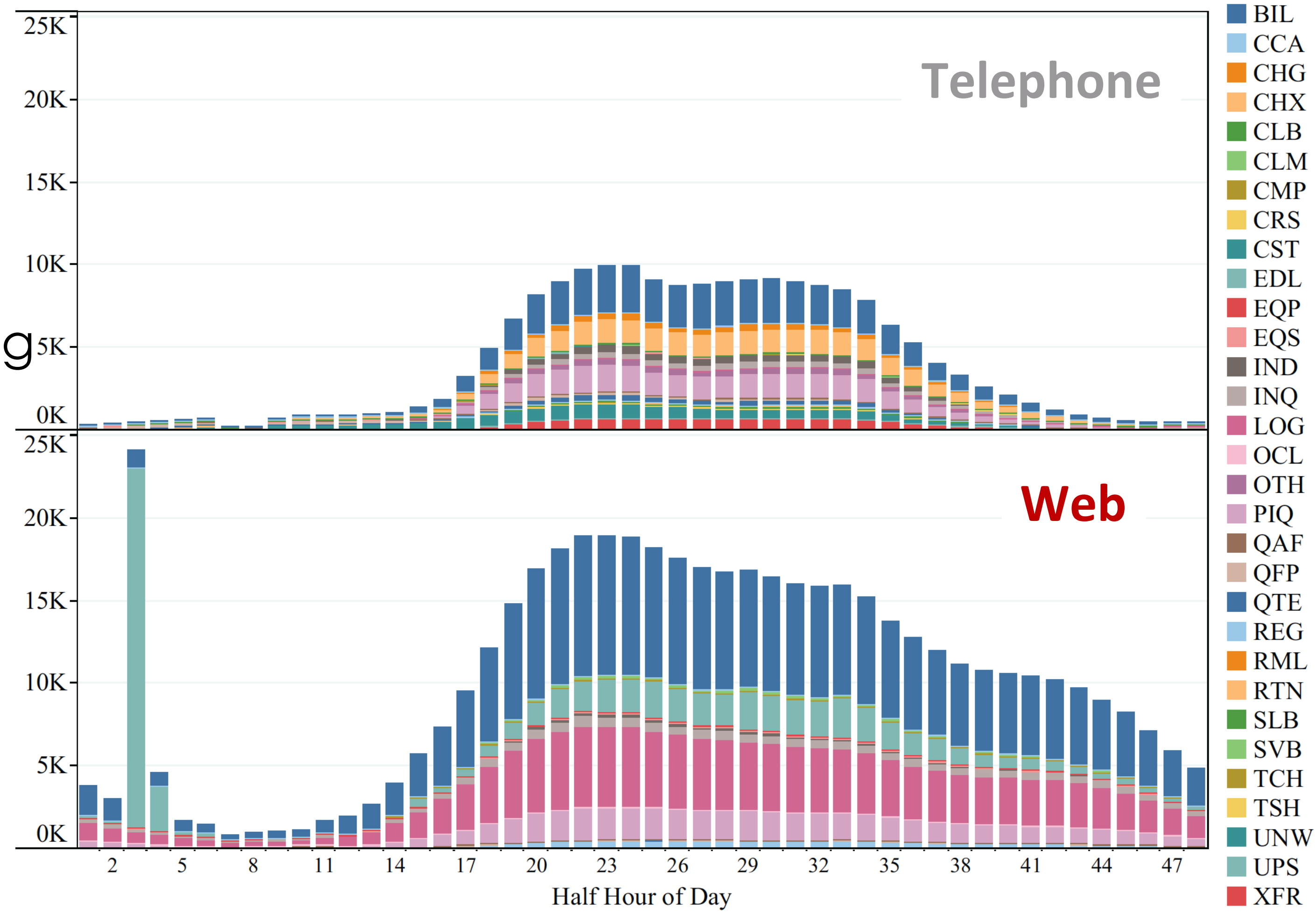
- Daily **Web** transaction volumes are consistently higher than the daily **Telephone** transaction volumes.
- Transaction volumes are higher during the weekdays, for both channels.
- Volume of daily Web transactions exhibits spikes on the first business day of each month (specified in Fig).
- Day-of-the-week effect in transaction volume, for both channels.

Contact Center Data

Averaged Hourly Volume of Transactions over a Day per Channel

Distribution of averaged hourly **Telephone** transaction volume has a convex decreasing tail and drops from 6 pm.

The **Web** transaction volume on average does not decrease during evenings.



Contact Center Data

Contact Reason Types (31 different categories):

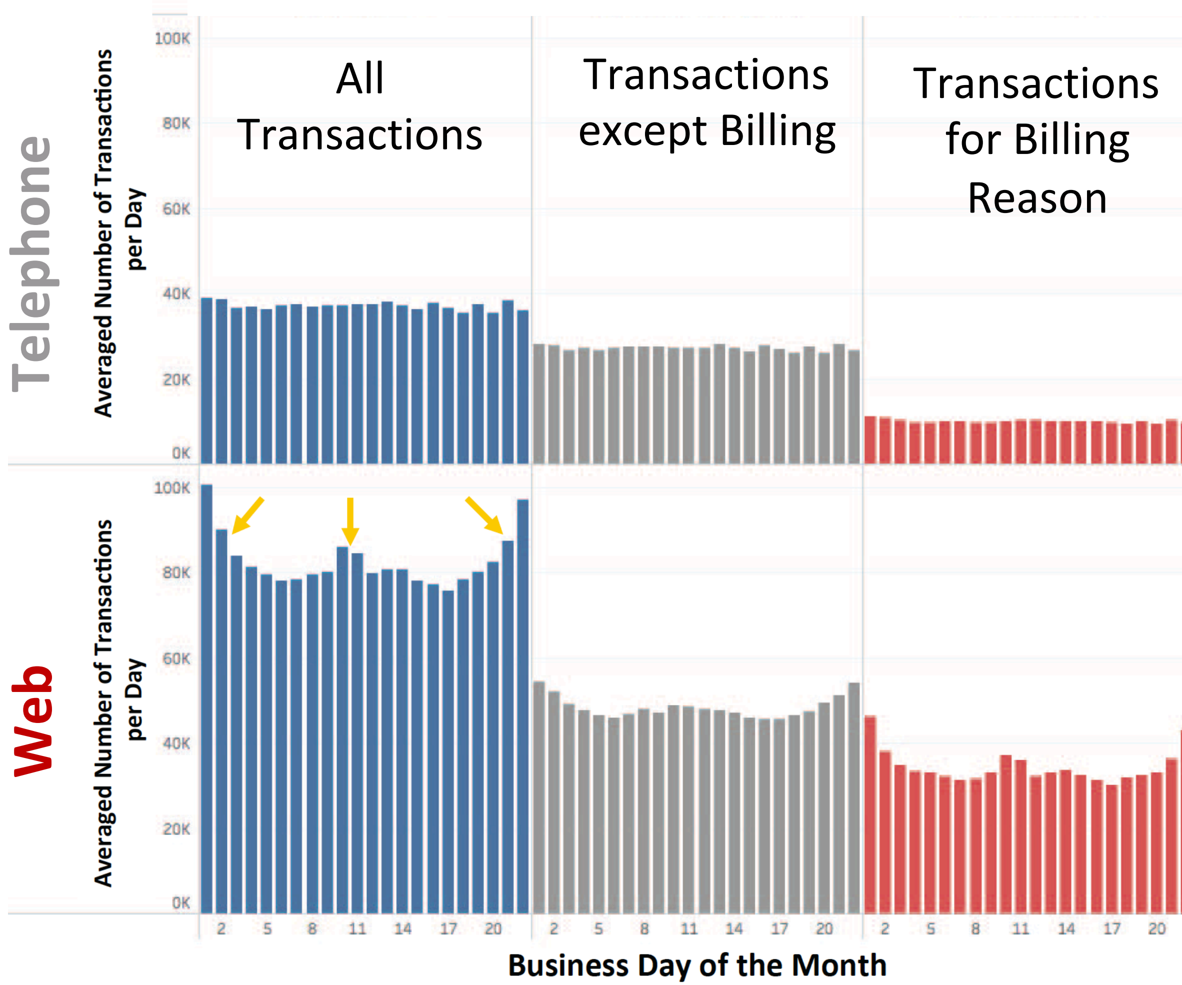
- Five categories make 81.3% of all transactions: Billing (37.9%), Login (17.7%), Policy Inquiry (11.2%), Electronic Message Delivery (10.2%), Policy Change (4.2%).
- **Web** Transactions are associated to 12 contact reasons.

  Four contact reasons, Billing (42.1%), Login (24.6%), Electronic Message Delivery (14.2%), Policy Inquiry (9.8%), constitute the reason for 90.7% of the Web transactions.
- **Telephone** Transactions are associated to 28 different contact reasons, including Billing (27.0%), Policy Change (15.1%), Policy Inquiry (14.9%), Underwriting (9.1%), Transfer (7.0%), Insurance Document (5.1%).

Contact Center Data

Averaged Number of Transactions per Day per Channel

Effect of billing days is prominent for Web transactions and with the contact reason Billing.



Individual Customer's Call Arrival Prediction

- **Goal:** to develop a feature-based predictive model to rolling forecast the occurrence of a call event by a policyholder over a set period of time ahead.
- Rolling Forecast Window: 30 days

$$y_i|\mathbf{x}_i := \begin{cases} 1 & \text{if policyholder } i \text{ contacts the company} \\ & \text{in the next 30 days via Telephone} \\ 0 & \text{otherwise} \end{cases}$$

Individual Customer's Call Arrival Prediction: Feature Modeling

Feature Class	Feature Subclass	# of Features
Customer Related Features	Date and Time	5
	Billing Cycle	3
	Contact Channel	5
	Participant Type	5
Reason Features	Contact Reason Type of Last Event	31
	Contact Reason Subtype of Last Event	81
Recency Features	Recent Contact in the Past 1, 7, 30 Days	3
	# of Days since the Last Contact	1*
Frequency Features	# of Past Events	1*
	# of Events in the Past Days	3*
	Average # of Days Between Events	1*
Cross-class Features	# of Days since Last Event per Channel	3*
	Cumulative # of Changes in Channel	1*
	# of Past Events per Channel	6*
	# of Past Events per Contact Reason	21*

Note. (*) indicates continuous features. All others are binary features.

(170 Features in total)

Individual Customer's Call Arrival Prediction: Feature Modeling

- **Date and Time:** Weekday, Holiday, period of the day of the last contact (0 am-8 am, 8 am-8 pm, 8 pm-0 am).
- **Contact Channel:**
  - Channel of the Last Contact,
  - Used multiple channels at least once in the past,
  - Whether policyholder used the same channel in the last contact as the channel in the exact one contact before the last contact, and the direction of the change
- **Billing Cycle:**
  - Last contact occurred on 1, 2, 10, 11, 21, 22 business day of month,
  - Last contact occurred on the 1, 2, 21, 22 business day of month,
  - Last contact occurred on the 10 or 11 business day of month

Individual Customer's Call Arrival Prediction: Feature Modeling

- **Recency Features:**
  - Whether there has been at least one (either Web or Telephone) contact in the past 1, 7, 30 days.
  - # of Days since the Last Contact.
- **Frequency Features:**
  - Total number of past contacts by the policyholder
  - # of contacts in the past 1, 7, and 30 days
  - average number of days between consecutive contacts by a policyholder
- **Cross-Class Features:**
  - Channel-Recency, Channel-Frequency,
  - Frequency-Recency-Contact Reason

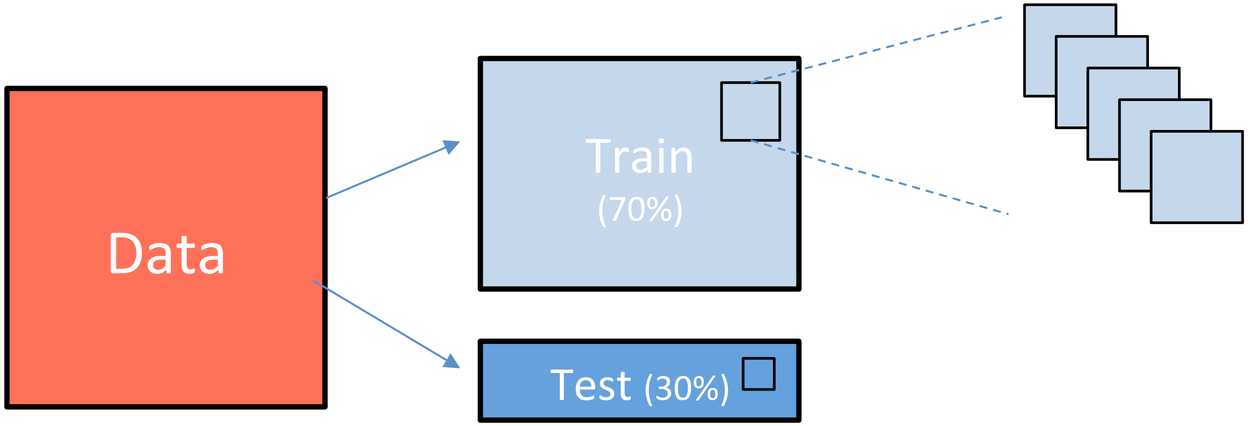
Individual Customer's Call Arrival Prediction: Methodology

- Feature selection becomes fundamental to reduce dimensionality, training time, to improve prediction performance.
- Scalable Data Analytics:

1. Mixed-effect Logistic Model

2. Lasso Method

3. Subsampling



Individual Customer's Call Arrival Prediction: Methodology

- **Response Model:** $\mathbb{E}\left(y_i \mid b_i\right) = g^{-1}\left(x_i^\top \boldsymbol{\beta} + b_i\right)$
- **Lasso Method:**

$$(\boldsymbol{\beta}^*, \sigma^*) = \underset{\boldsymbol{\beta},\sigma}{\arg\min}\left\{-\log\left(\mathcal{L}(\boldsymbol{\beta}, \sigma)\right) + \lambda \left\|\boldsymbol{\beta}\right\|_1\right\}$$

- **Minimizing the Bayesian Information Criterion:**

$$\lambda^* = \underset{\lambda}{\arg\min}\left\{-2\log\left(\mathcal{L}(\boldsymbol{\beta}(\lambda), \sigma(\lambda))\right) + |\mathcal{A}(\lambda)| \log \ N\right\}$$

$$\mathcal{A}(\lambda) \ := \ \{k : \beta_k(\lambda) \neq 0\} \qquad N := \text{ Number of Contacts}$$

- Kleiner, Talwalkar, Sarkar, Jordan (2014): S = 435 training subsample datasets of size M = 9,080 are considered.

Individual Customer's Call Arrival Prediction: Estimation and Results

- 14 Features selected in more than 50% of sampled datasets.

Rank	Feature	Avg. Coef.	S.E.	p-value
-	(Intercept)	-4.3860	0.1155	1.14E-140
1	Number of Contact Reason Type - e-delivery (EDL) in the Last 30 Days	0.2122	0.0022	1.07E-294
2	Days Since Last Event	-0.4530	0.0038	$<$1.0E-400
3	Days Since Last Telephone Event	-0.3675	0.0036	1.90E-307
4	Number of Telephone Events in the Last 30 Days	0.1041	0.0024	2.23E-159
5	Contact Reason Type of Last Event - Login (LOG)	-0.3514	0.0063	2.54E-201
6	Number of Telephone Events in the Last 7 Days	0.0777	0.0043	6.82E-55
7	Participant Type Last Event - Customer	0.1382	0.0025	4.51E-202
8	Contact Reason Type of Last Event - Quote Acceptance Form (QAF)	0.1633	0.0187	5.65E-17
9	Contact Reason Type of Last Event - Billing (BIL)	-0.1936	0.0027	8.73E-247
10	Number of Contact Reason Type - Underwriting (UNW) in the Last 30 Days	0.0713	0.0035	1.74E-64
11	Contact Reason Type of Last Event - eQuote Acceptance Package (EQP)	0.0629	0.0074	2.25E-16
12	Contact Reason Subtype of Last Event - Quote Acceptance Package (QAP)	0.0673	0.0018	1.17E-138
13	Number of Contact Reason Type - Policy Inquiry (PIQ) in the Last 30 Days	0.0385	0.0021	1.67E-57
14	Number of Telephone Events in the Last 1 Day	0.0459	0.0034	1.19E-34

Note. All features are statistically significant at 0.001 level.

- Customer related (1), Recency related (1), Call Reason related(5),
- Cross-class features (7): six of which are Frequency Related.

Individual Customer's Call Arrival Prediction: Estimation and Results

- *e-delivery* in Last 30 Days has a significant positive impact on probability of a call arrival: effect of the policyholder's Web activities on the probability of his future calls.
- Negative coefficient of *Days Since the Last Event:* the more recent a policyholder contacted the company, the higher the probability that he will make a telephone query in the next 30 days.
- Positive influence of *Contact Reason of Last Event – QAF*, *Contact Reason of Last Event - EQP, Contact Reason of Last Event – QAP*: suggests that a follow-up contact with customers with questions on new contract will occur.

Conclusions

- Analyzed effectiveness of characteristics of a policyholder and his previous **Web-Telephone** contacts and their reasons on the probability that he will call in the next 30 days.
  - Policyholder-Level prediction
  - Massive Data Set (35 million contacts)
  - Rich Set of Features
- Found evidence of relevance of recent Web Activities.
- Recency & Frequency significantly increases probability of call.
- Modeling approach with the set of selected features enables businesses to identify opportunities to act proactively in an attempt to solve eventual problems of those customers who are more likely to call back in the short term.

STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

stevens.edu

Thank you!