Predicting Individual-Level Call Arrival from Online Account Customer Activity

Somayeh Moazeni

Stevens Institute of Technology

Hoboken, NJ, USA

smoazeni@stevens.edu

2018 IEEE INTERNATIONAL CONFERENCE ON BIG DATA

SEATTLE, WA, USA

accenture

DEC. 12, 2018

STEVENS INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY

Big Data and Contact Centers

- Customer relationship management and contact center strategies involve predicting customers' behaviors.
- Pivotal prediction problem:

Consumer Call Arrivals

- Complexity:

Individual-Level Predictions

Goal: Leverage customer **online account activities** to predict future telephone queries by an individual customer to the firm's contact centers.

Outline

- Related Literature
- Individual-Level Call Arrival Prediction
- Contact Center Data
- Feature Engineering
- Artificial Neural Network Models
    - Performance Evaluation Metrics
    - Training Runtime
- Results and Discussion
    - Actual and Predicted Call Rates
    - Individual-Level Prediction
- Conclusion

Related Literature

Call Arrival Prediction

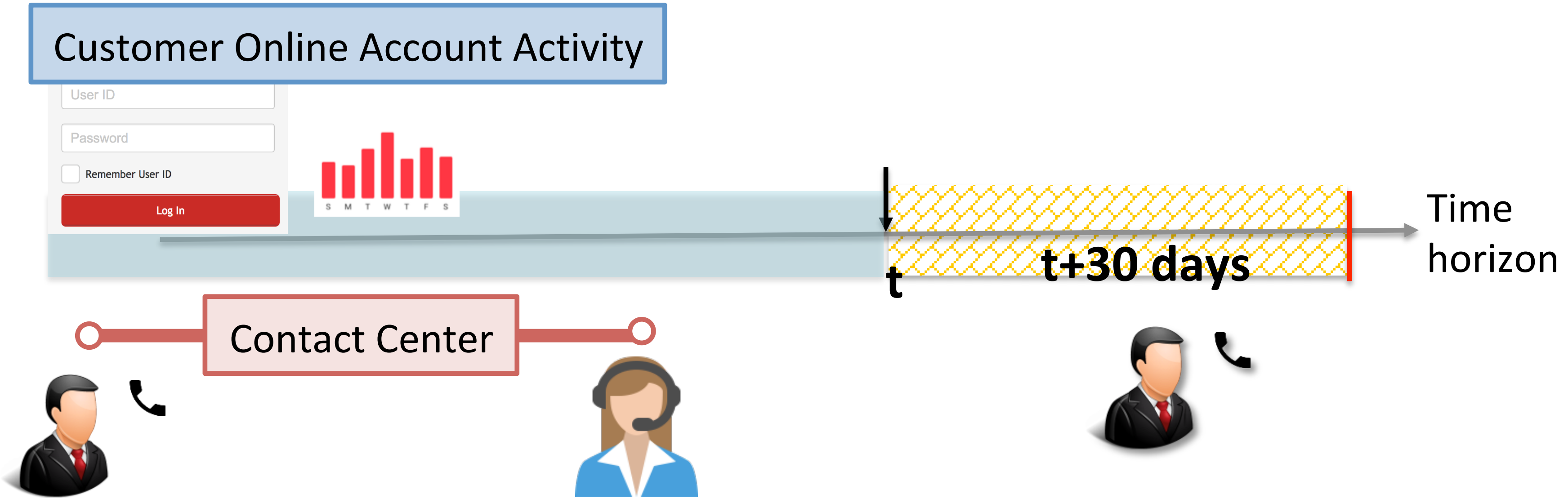
- Literature focused on estimating **arrival rate of calls** using **historical telephone queries.**
- Our study:
  - Customer-level Predictions, Web data

Prediction using Web Activity Data

- Data collected from web search and web activities has demonstrated to have the capability to predict certain behaviors public health and marketing.

Individual-Level Call Arrival Prediction

- **Prediction Problem:** to forecast whether an individual customer is going to call the firm's contact centers within the next 30 days, given the characteristics of the customer and the customer's past interactions with the firm by either telephone or firm's online account.



Contact Center Data

- Data provided by a major U.S. insurance firm
- Jan. 1, 2015 to Dec. 31, 2015.
- 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 contact transactions
Policy ID Attributes	Attributes to identify a policyholder
Contact Channel	medium used for the contact
Event Timestamp	Date and time of the contact
Participant Type	Role of the participant who contacts
Contact Reason Type (31)	High-level reason for the contact
Contact Reason Subtype (81)	Detailed reason for the contact

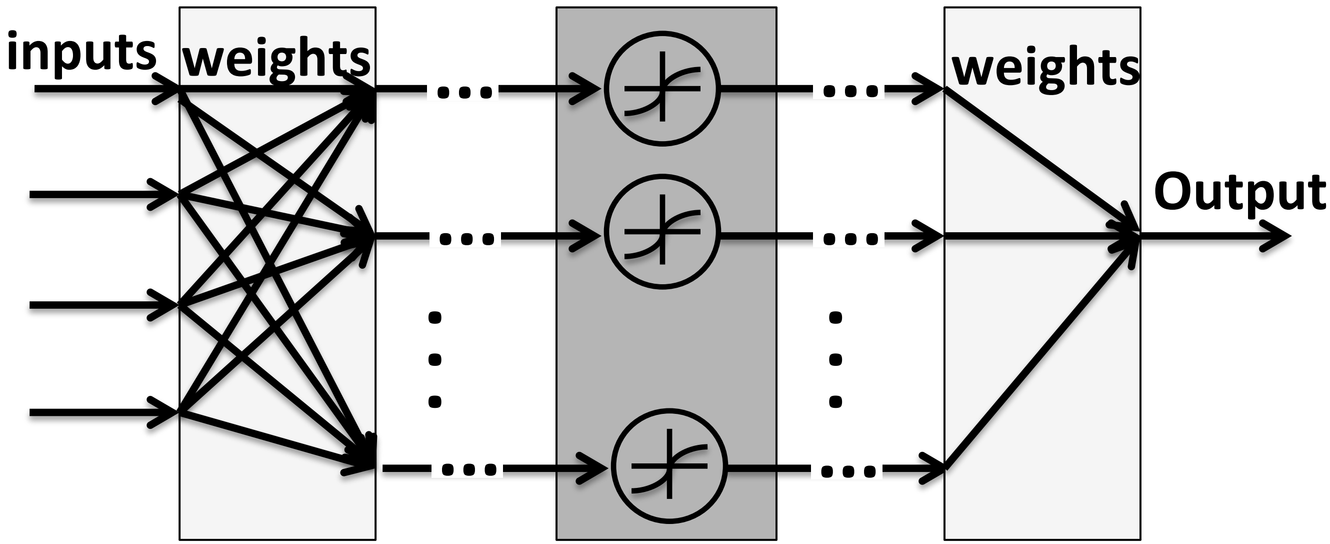
Feature Engineering

Feature Class	Feature Subclass	# of Feature
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 non-binary features. All others are binary features.

(170 Features in total)

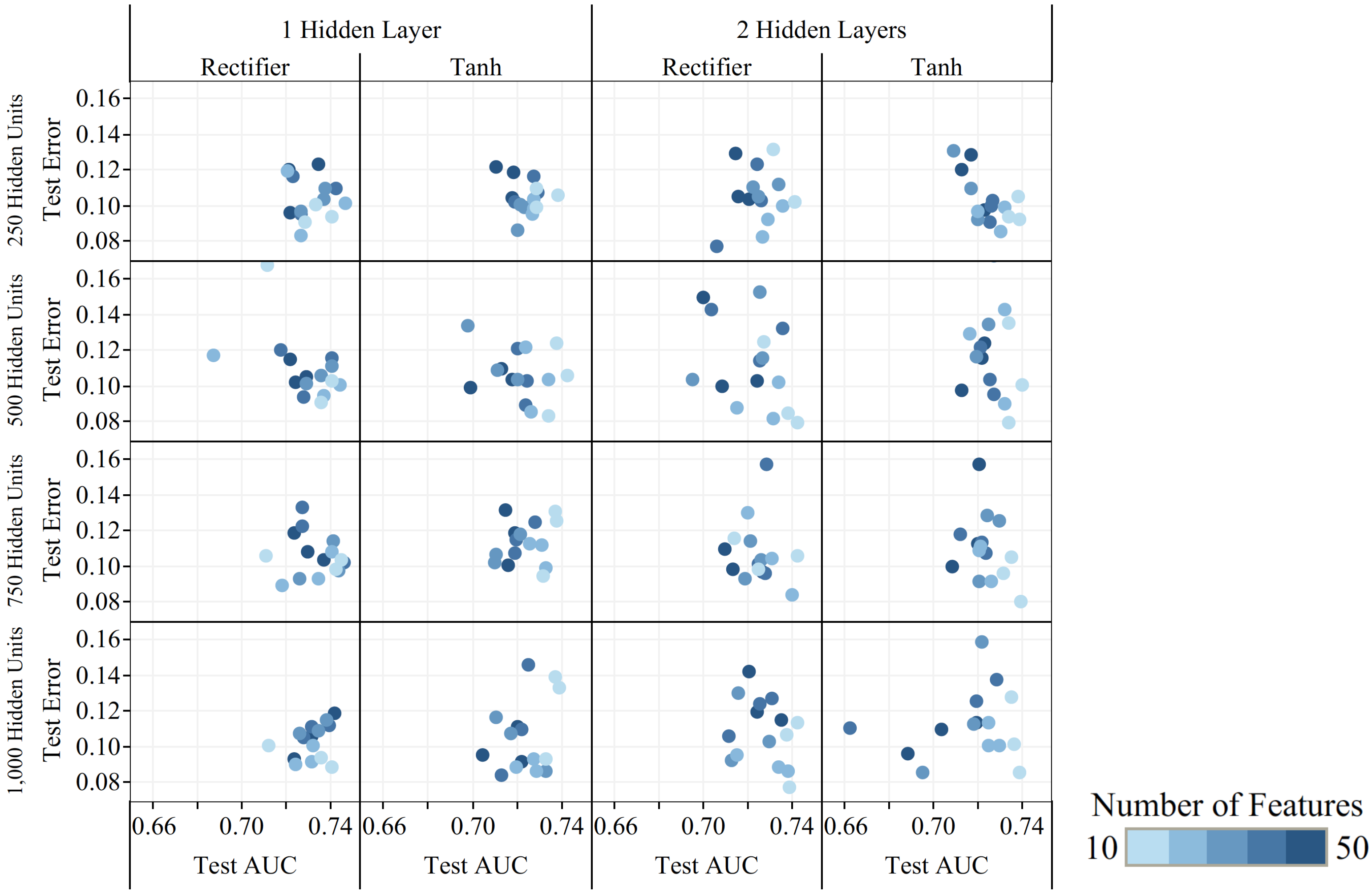
Artificial Neural Network Models



- ANN is a composition of functions $h = h_m \circ h_{m-1} \circ \cdots \circ h_1$
- Investigate models with different hidden layers, hidden units, activation functions (hyperbolic tangent, Rectifier), number of input features, regularization parameters.
- Performance evaluation metrics: misclassification error, AUC

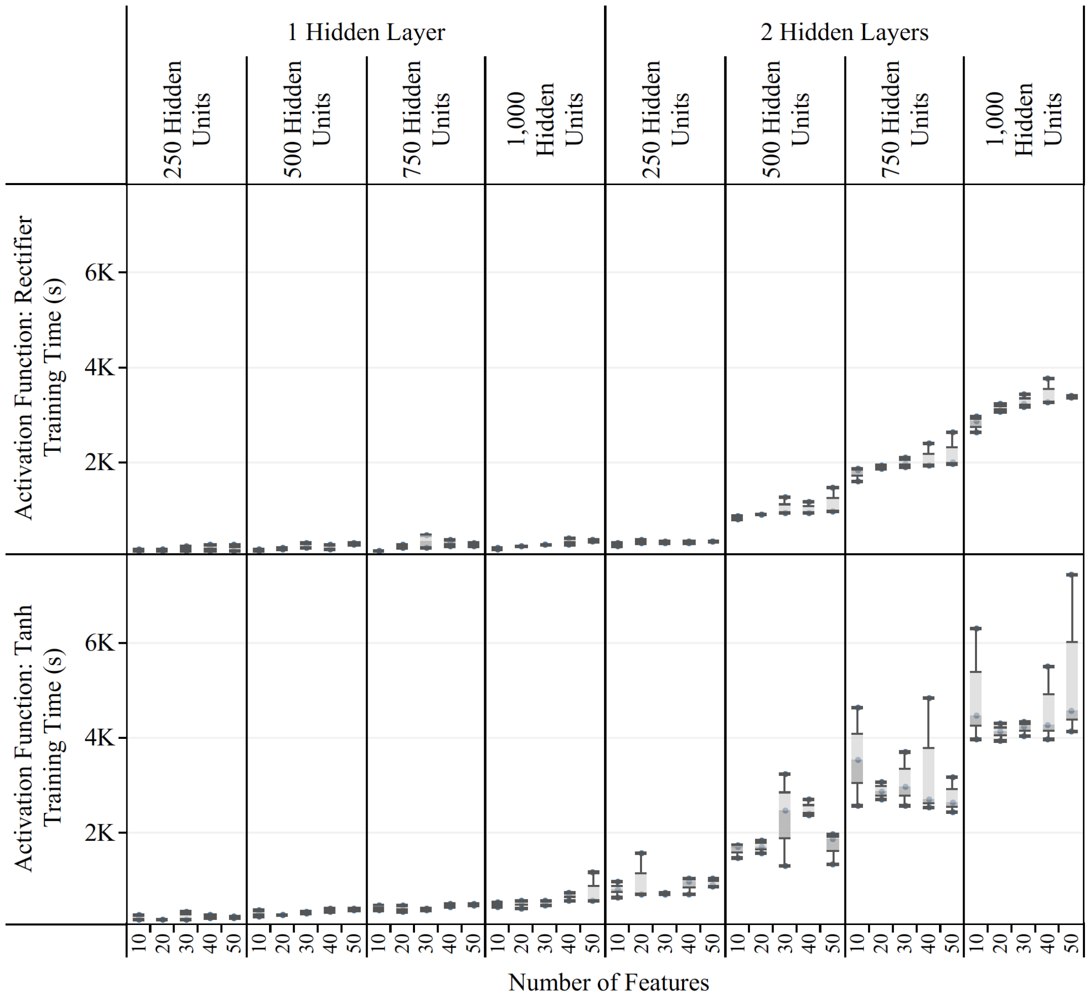
Artificial Neural Network Models

- Performance metrics for different ANN Architectures



Artificial Neural Network Models

- Training time for different ANN Architectures



Results and Discussion

- **Actual and Predicted Call Rates**

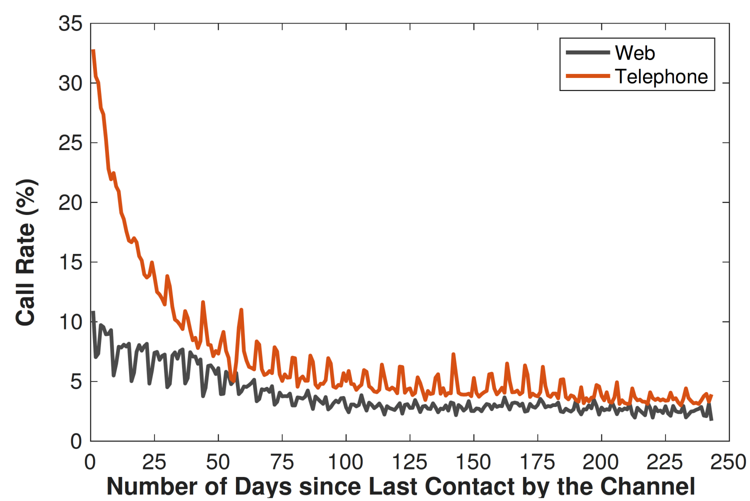
set of policyholders who make a telephone call in the forecast window

$$\text{Call Rate}(\mathbf{x}^P) := 100 \times \frac{|\mathcal{A}(\mathbf{x}^P) \cap \mathcal{Y}|}{|\mathcal{A}(\mathbf{x}^P)|}$$

feature vector

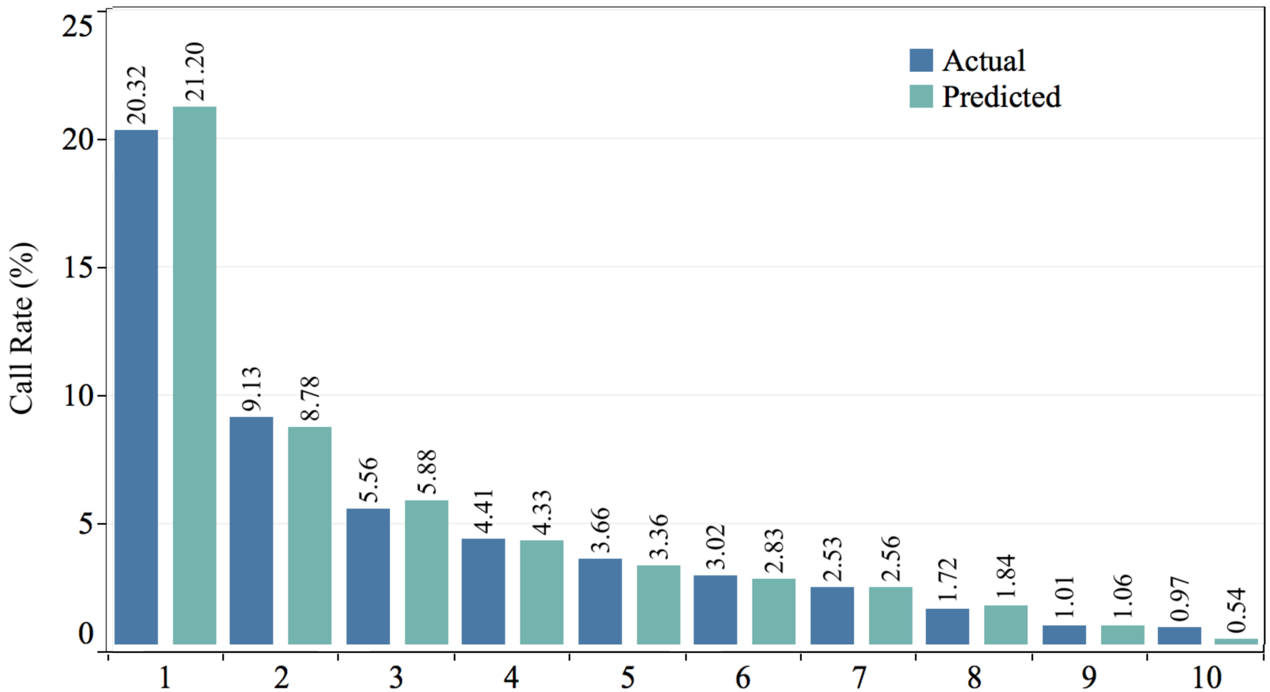
set of policyholders characterized by $\mathbf{x}^P$

- Let $\mathbf{x}^P$ specify a policyholder whose last contact occurred *T* days before the snapshot date via the contact channel c.



Results and Discussion

- **Individual-Level Prediction**
  - To assess the model at the individual-level, we (i) introduce a segmentation for policyholders by computing the call likelihood for each policyholder in the out-of-sample dataset using the trained model, (ii) rank the policyholders based on predicted call probabilities.



Population Deciles Sorted by the Predicted Call Likelihood

Results and Discussion

- **Individual-Level Prediction**
  - We compare the percentage of calling policyholders identified by the model who actually called in the forecast window.
  - A calling policyholder refers to a policyholder who is predicted to make a telephone call according to the model.

Top percentile of policyholders based on their calling probabilities [%]	Correctly identified calling policyholders [%]
20	54.9
40	74.6
60	87.7
80	96.0
100	100

Identification of Calling Policyholders

- For the top 60% of policyholders identified as calling policyholders by our model, 87.7% actually called in the (out-of-sample) forecast window.
- These results demonstrate our model's ability to effectively identify calling policyholders.

Conclusion

- Predict whether a policyholder will call the customer support centers of an insurance firm within the next 30 days:
  - Individual-level prediction
  - Leveraging firm's online account customer activities
  - Rich set of input features (e.g., detailed contact reasons, recency and frequency of past contacts)
- Transaction-level data provided by a major US insurance firm.
- Model provides accurate predictions for aggregate telephone query volumes.
- Model is capable of detecting with high accuracy the customers who are likely to call.
- Effect of frequency differs for different contact reasons and channels.

Thank you!

Questions and Comments:

Somayeh Moazeni

smoazeni@stevens.edu