Deep Neural Networks for Multi-Touch Attribution

e-book by: Express Analytics

Introduction

Online retail has become an important part of our society. Along with online shopping, advertisements, too, are now an integral part of a marketing and sales strategy to attract new customers. For a retailer, it is now crucial to know which touchpoints and advertisements are effective. Millions of customers are exposed to advertisements over a wide range of digital channels. Organizations spend big bucks on advertising and marketing like email advertising, and on Google Adwords, Instagram, Facebook and Twitter. They also bid on results of search engines (e.g. Google and Bing) for certain high-ranking keywords.

The decision of allocating budget across channels is commensurate with their respective impact in your market. Marketers track all customer encounters with these channels. With deep learning, however, there's been a paradigm shift in data science. Millions of customers are exposed to advertisements over a wide range of digital channels. These interactions leave behind patterns buried deep inside, waiting for marketers to uncover them. Deep learning has proved to be a reliable technique to extract these patterns and predict outcomes from large datasets. Also, for a targeted campaign and for the allocation of the right budget, a marketing team must be able to confirm the influence of different marketing channels on audiences. Given the success of deep learning models on large datasets, implementing artificial neural networks seems to be the most promising approach.

In this e-book, we talk about a novel attribution algorithm based on deep learning that Express Analytics implemented in case of a client in order to assess the impact of each advertising and marketing channel.

Disclaimer

We have been working with this client and have implemented this and other models using their customer data. For reasons of privacy and confidentiality, we are unable to disclose the details such as the name of the client, the names of its customers, or other such sensitive data. For the reader to get the picture, we've replaced real instances with mock data. However, the performance (AUC metric) quoted corresponds to the actual data.

There is a gap between the data any marketer has, and the information needed for making a sound business decision. A wide variety of parochial rule-based models and data-driven models (e.g. Markov model) have been widely adopted but do not address all of channel interaction, time dependency, and user characteristics.

A deep neural network single-handedly closes these gaps. The only assumption it makes is that only the data speaks for itself. In fact, the model welcomes user context information - browser used by the prospective customer, device type, device name; any channel or device from which we can see a pattern emerging. By looking for patterns in the data, it gives a fair way to assign credit for sales to channels in conversion paths.

The illustration below shows how the data is "pre-processed" before being fed to the neural network.

cookie	time	conversion_value	channel	device_type	device_name	os	browser
000dfkffd69aldfa	2018-07-03T13:02:11Z	0	Instagram	Mobile	ONEPLUS 7T	Android	Not set
000dfkffd69aldfa	2018-07-17T19:15:07Z	0	Online Display	Mobile	ONEPLUS 7T	Android	Chrome
000dfkffd69aldfa	2018-07-24T15:51:46Z	0	Facebook	Mobile	ONEPLUS 7T	Android	Not set
000dfkffd69aldfa	2018-07-29T07:44:51Z	0	Online Display	Mobile	ONEPLUS 7T	Android	Chrome
000qirngd89hiopf	2018-07-03T09:44:57Z	0	Paid Search	Desktop	(NOT SET)	Windows	Mozilla Firefox
000qirngd89hiopf	2018-07-03T23:36:49Z	0	Paid Search	Desktop	(NOT SET)	Windows	Mozilla Firefox
000qirngd89hiopf	2018-07-10T22:24:41Z	0	Paid Search	Desktop	(NOT SET)	Windows	Mozilla Firefox
000qirngd89hiopf	2018-07-10T22:24:44Z	0	Online Video	Desktop	(NOT SET)	Windows	Mozilla Firefox
000qirngd89hiopf	2018-07-11T21:21:46Z	0	Paid Search	Desktop	(NOT SET)	Windows	Mozilla Firefox
000qirngd89hiopf	2018-07-11T22:19:53Z	0	Online Video	Desktop	(NOT SET)	Windows	Mozilla Firefox
010vifapo72pxikh	2018-07-08T10:48:21Z	0	Facebook	Mobile	IPHONE XS	IOS	SAFARI
010vifapo72pxikh	2018-07-09T13:06:42Z	0	Instagram	Mobile	IPHONE XS	IOS	SAFARI
010vifapo72pxikh	2018-07-10T13:02:36Z	0	Instagram	Mobile	IPHONE XS	IOS	SAFARI
010vifapo72pxikh	2018-07-10T14:04:14Z	0	Online Display	Mobile	IPHONE XS	IOS	SAFARI
010vifapo72pxikh	2018-07-12T10:34:38Z	100	Paid Search	Mobile	IPHONE XS	IOS	SAFARI
214etfipd89ppsmt	2018-07-09T16:57:18Z	0	Instagram	Tablet	LENOVO E10	Android	Opera
214etfipd89ppsmt	2018-07-17T16:00:58Z	0	Facebook	Tablet	LENOVO E10	Android	Opera
214etfipd89ppsmt	2018-07-17T16:01:44Z	0	Facebook	Tablet	LENOVO E10	Android	Opera
214etfipd89ppsmt	2018-07-18T17:17:24Z	250	Instagram	Tablet	LENOVO E10	Android	Opera

channel	number
Instagram	0
Online Display	1
Online Video	2
Facebook	3
Paid Search	4

device_type	number
mobile	0
desktop	1
tablet	2

- For each channel in the sequence, we calculate how old each encounter was, compared to the last encounter. We normalize this array at the end. 1 corresponds to the oldest touchpoint and 0 to the latest.
- The channels, device name, device type, operating system and browsers are categorical variables, and we assign indices to them.
- These indices are fed into an embedding layer whose weights are indexed out as learnable vector representations.

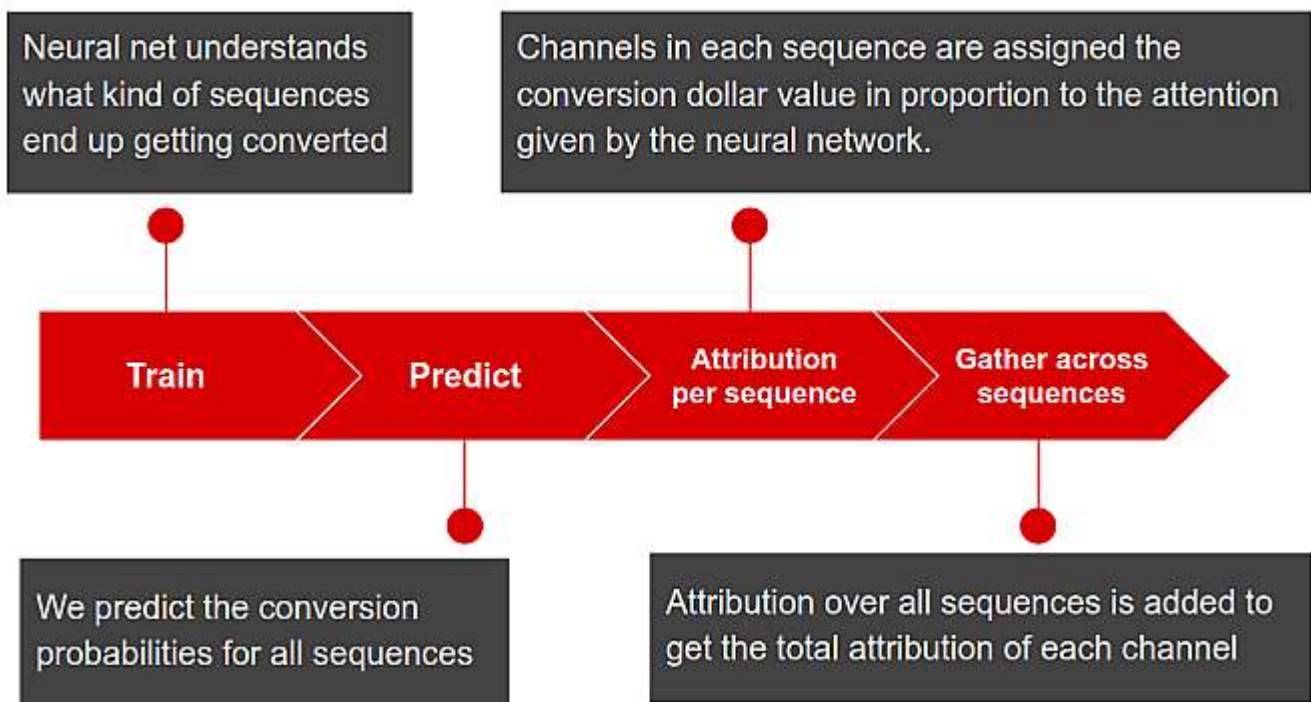## Modeling activity of unknown users

It is important that we mention here that the data deals with not only the users who already have an account with an online store but also the customers who anonymously visit its website. Since our neural network forms its opinion about channels using both converted and unconverted sequences, adding activity of unknown users prevents the model from having biases due to misrepresentation.

Foot traffic attribution

Foot Traffic Attribution connects ad exposure to real world behavior to quantify the impact of digital and out-of-home advertising on in-store visits.

It is possible for several users to have bought products from an offline store, or perhaps because of their online activity. The data collected also includes most purchases at the store. Thus, we can be confident that the attribution values extracted are sourced from all the data our client can possibly find.

Neural net understands what kind of sequences end up getting converted

Channels in each sequence are assigned the conversion dollar value in proportion to the attention given by the neural network.

Train ⟩ Predict ⟩ Attribution per sequence ⟩ Gather across sequences

We predict the conversion probabilities for all sequences

Attribution over all sequences is added to get the total attribution of each channel

The premise is performance

A deep neural network, in a supervised learning fashion, learns to predict if a series of encounters with touchpoints leads to conversion. By doing this, the model inevitably ends up having a deep understanding the effects of dynamic interaction between channels.

How do we know whether the network has captured such patterns in the dataset?

For this case of "imbalanced data", we chose the AUC metric.

- With touchpoints appearing in a large number sequences, the neural network can infer from both converted and unconverted sequences. For example, if the channel display ads occur in equal amounts in both converted and unconverted sequences, the neural network can never amount to a good classifier if it paid more attention to the display ads in the sequence.
- Thus, the attention values given by the deep model can be directly used as a fractional score, as it serves as the contribution of each touchpoint after accounting for the interaction between each other.
- If the model is not giving appropriate weightage to the touchpoints in proportion to their impact, it will not be able to separate out converted (positive) and unconverted (negative) sequences.
- If the neural network is able to achieve a good test ROAUC on the dataset, we can be sure that the model has captured enough patterns in the dataset. The attributions given by the neural network are that much in agreement with the ground truth.
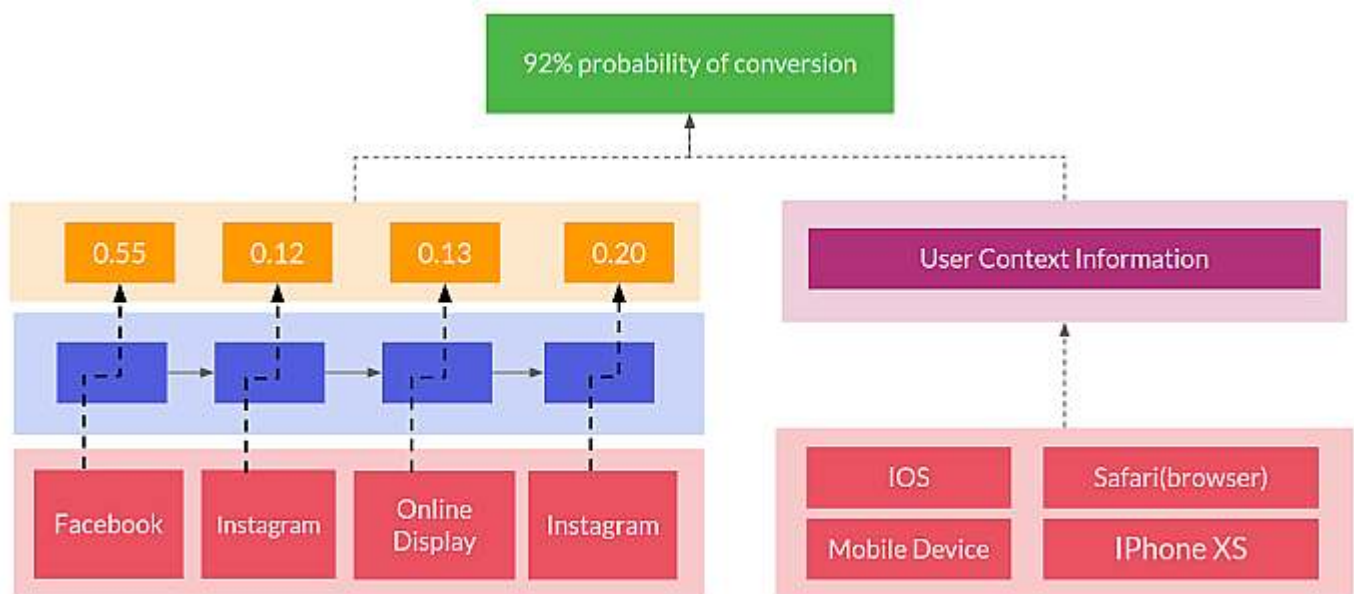
The above is an illustration of the neural network for an example sequence of 4 touch-points that a user encountered.

LSTM layer (blue layer)

LSTMs are well known and widely used for processing sequential data. The embeddings of channels are fed here. The output would be a numerical representation of the effect of the impressions the user has come across. For the above sequence consisting of 4 encounters of a user with our channels, we get 4 vectors as output.

Attention layer (yellow layer)

$$v_t = tanh(W_v h_t + b_v)$$

$$a_t = \frac{exp(v_t^T u)}{\sum_t exp(v_t^T u)}$$

In a sequence of observations of touchpoints, the same touchpoint may be differentially important at different time locations and at different frequencies of occurrence. This attention mechanism that lets the model pay appropriate emphasis on individual touchpoints when constructing the representation of the customer path.

It is important that the attention scores (amount of emphasis) are calculated during the forward pass in the network. The model has the capacity to give channel 'A' a score of 50% in the sequence [A,B,C] whereas 20% in the sequence [C, A, D]

User context information (maroon layer)

The neural network welcomes information specific to a user such as device type, device name, operating system. This information opens up possibilities for finding patterns in the data, enabling a much more informed prediction about the sequence conversion. This information is numerically captured in the dense layer.

Output layer (green layer)

The probability of conversion is predicted using the output from the attention layer and the dense layer.

Attention layer with time decay

$$a_t \quad = \quad \frac{exp(v_t^T u - \lambda T_t)}{\sum_t exp(v_t^T u - \lambda T_t)}$$

This can be used instead of the default attention layer. This uses time relative to the last channel in the sequence to penalize attention/attribution.

The decay parameter lambda is learnt from the training data. It is clear that lambda must be positive. In the case when everything else in the sequence is quite the same, the network must provide more attribution to the last few channels in the sequence because it converted the customer.

Area under ROC

This section tabulates the performance on and the attribution values extracted from mock data. This is useful only for analysis of models.

Model	AUC
Attention Neural Network	0.89
Attention Neural Network with time decay	0.96
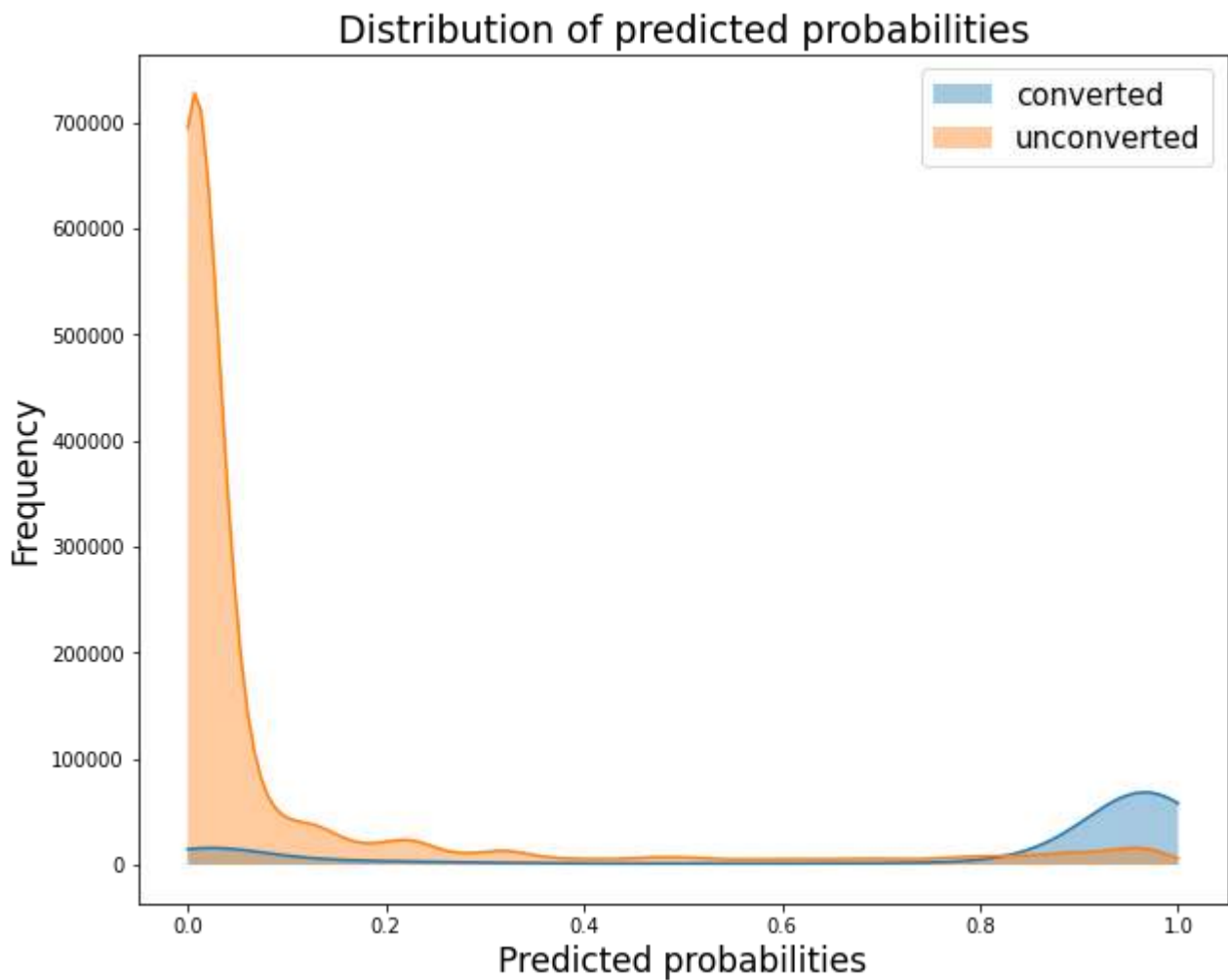
Distribution of predicted probabilities

This plot shows the distribution of probabilities predicted by the neural network. For unconverted sequences, we can say most of the predictions were less than 0.2 whereas those for the converted sequences were greater than 0.8. This illustrates the power of the network in separating converted and unconverted sequences.
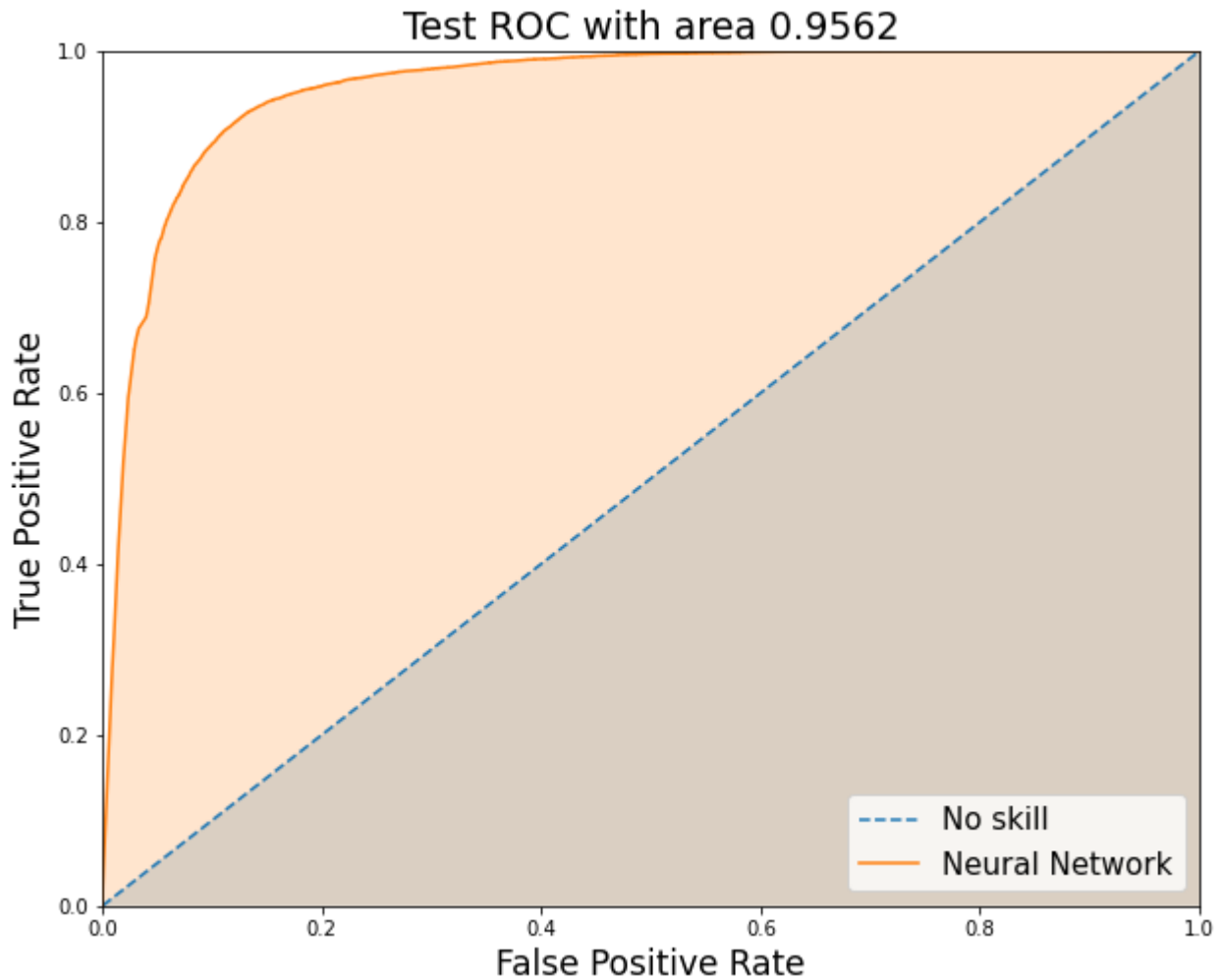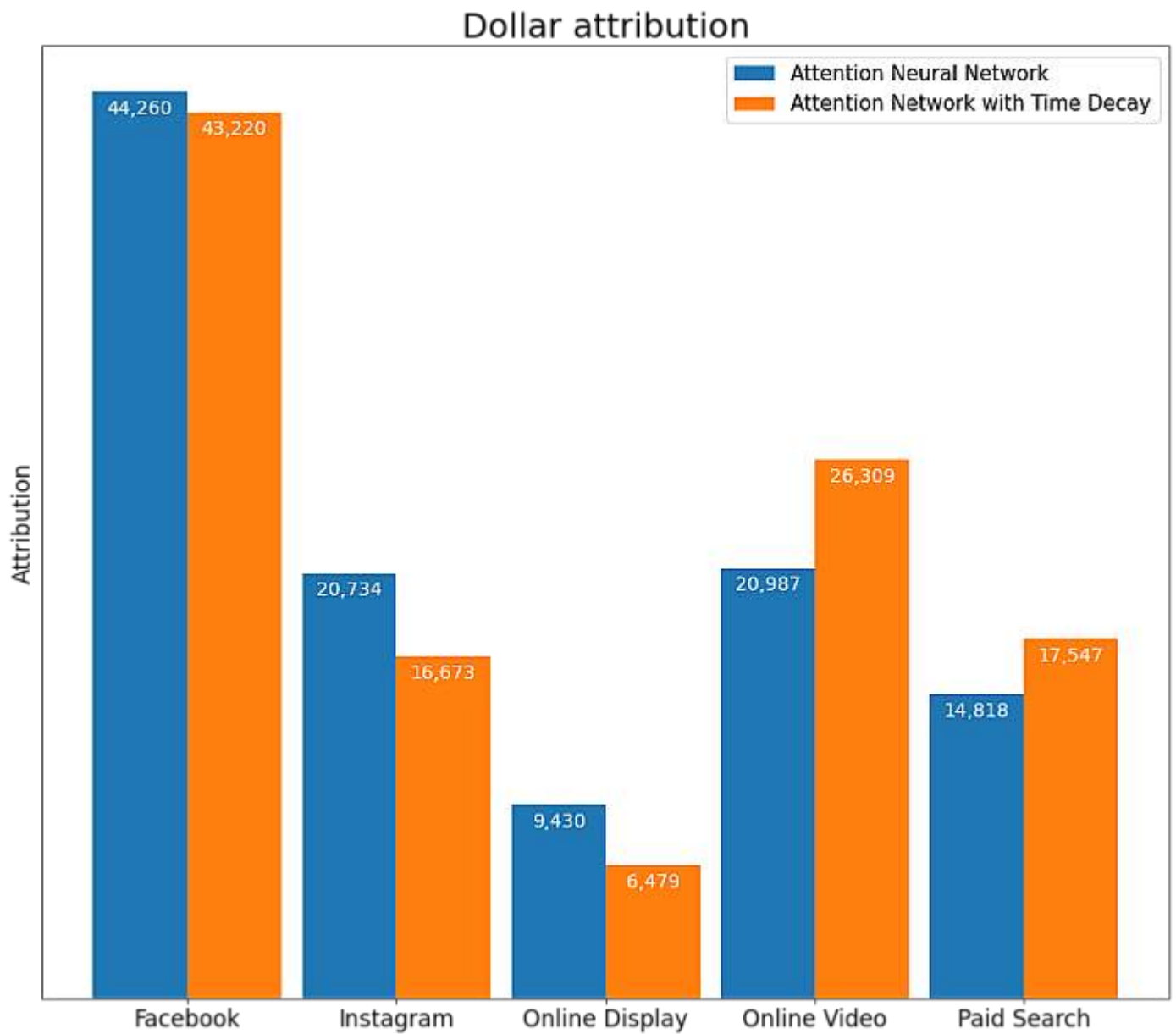

This shows the ROC of the Time Decay Attention Neural Network.

Disclaimer: The dollar attribution values herein are merely for the purpose of the presentation and do not necessarily reflect the views of Express Analytics on your browser data.

Dollar attribution

Channel	Attention NN	Attention NN with Time decay
Facebook	44,260	43,220
Instagram	20,734	16,673
Online Display	9,430	6,479
Online Video	20,987	26,309
Paid Search	14,818	17,547

References

- Research paper implemented https://arxiv.org/abs/1809.02230
- Mock data obtained from this blog on markov model
- Foot traffic attribution: https://www.localpagepop.com/lba/foot-traffic-attribution/
- Colah's article on LSTMs : https://colah.github.io/posts/2015-08-Understanding-LSTMs/
- Jay Alammar's article on transformers and attention: http://jalammar.github.io/illustrated-transformer/

About Express Analytics

Express Analytics offers a slew of services such as retail analytics, business intelligence analytics solutions and customer analytics. Our customer data platform "Oyster" is a world-class data unifying platform for B2C and B2B businesses. To know more:

Get In Touch