Application of Event Sequence CNN Framework on Customer Churn Prediction

Cheng Ji

Agenda

- Background
- Problem Statement
- Data Preparation
- Modeling approach
- Summary

Background

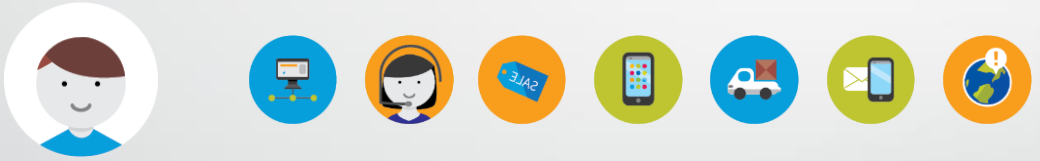
- Client: One of the largest wireless carriers in US
- Current solution: Logistic Regression (AUC=0.69)

Background

- **Motivation:**
- Current feature pool has around 900 variables, but it still **lacks of variety**. We have numerical variables covering mainly bill, payment, voice/data usage and some account/plan info.
- There is one data we have access but yet to use: **Memo** data which records a lot of **events** happened to the customers with timestamp.
- **Question**:
- Can we do something with this data to improve our model performance?

Problem Statement:
Can we use Memo data to predict churn?

Historical events in the past X months



What's his **next** action?
- Churn, AAL or HSU?

Data Preparation

- What does the data look like?

ACCOUNT_ID	MEMO_DT	MEMO_TYPE_CD	ACCESS_METHOD_ID	MEMO_SYSTEM_TXT	MEMO_SOURCE_CD	CONV_RUN_NBR	MEMO_PRODUCT_TYPE_CD
950102715	7/5/2018	OBU		General			
957023041	7/5/2018	CFPM		Account accessed by System User ID - emallar6 on 07/05/2018 02:45:19, no changes made.			
963595765	7/5/2018	JPA3	4704499857	iPhone upgrade program SOC has been added to Subscriber 4704499857.			
958567564	7/5/2018	EL01	0				
944339307	7/5/2018	563		BAN SOC updated. SOC: UNLMSGF. Expiration date: 06/14/2018.			
510680036	7/5/2018	200	5204148840	EPSI transaction for activity <Add MA SOC> was processed and ended in: Success. (transaction_id=100000424			
759988760	7/5/2018	2115		Credit Policy manually changed: Tenure Upgrade performed. Approved Subscribers changed to 11 from <11>.			
406318956	7/5/2018	RLPR	9158200665	As of 07/05/2018 you have 20.00 GB available in your T-Mobile Data Stash in addition to your monthly data p			
952895859	7/5/2018	GEN	7027470721				
960033238	7/5/2018	MPYM		Payment Received: Amt = $71.00; Pymt Type = CREDIT CARD.			
917471057	7/5/2018	2115		Credit Policy manually changed: Credit Score to 940 from 508. Credit Class is B, Required Deposit is 0,Approve			
949328860	7/5/2018	PAYA		Pymt Auth. Pymt Deposit Date: 20180705. Amount: 0.0. Auth Via: POS. Code: Yes. Date: 20180705. Time: 13:3			
956883527	7/5/2018	UPBS		Decision Engine. Old behavior score: 511, new behavior score: 507, old behavior score type: C1, new behavio			
939874798	7/5/2018	PAYA		Pymt Auth. Pymt Deposit Date: 180705. Amount: 138.54. Auth Via: ESERVICE. Code: Yes. Date: 180705. Time:			
955780311	7/5/2018	VRFY		The details were verified by user: 148979 on 07/05/2018			

Data Preparation

- What does the data look like?

MEMO	SYSTEM_TEXT_TMPLT_DESC	cnt
CFPM	Account accessed by %s - %s on %s, no changes made.	19,758,240
VRFY	The details were verified by user: %s on %s	16,687,089
UPBS	Decision Engine. Old behavior score: %s, new behavior score: %s, old behavior score type: %s, new behavior score type: %s, date: %s.	7,446,028
MPYM	Payment Received: Amt = $%s; Pymt Type = %s.	7,011,645
PAYA	Pymt Auth. Pymt Deposit Date: %s. Amount: $%s. Auth Via. %s. Code: %s. Date: %s. Time %s.	6,178,766
1023	Subscriber level SOC created. SOC: %s. Effective: %s.	4,619,356
RLPR	As of %s you have %s%s available in your T-Mobile Data Stash in addition to your monthly data plan. Track usage at %s%s%s%s%s%s.	4,134,416
522	BAN level SOC created. SOC: %s. Effective date: %s.	3,659,755
1024	Subscriber level SOC expired. SOC: %s. Expiration date: %s.	3,478,735
2115	Credit Policy manually changed: Credit Class to %s from %s, Required Deposit to %s from %s, Approved Subscribers to %s from %s,Behavior score to %s from %s.	2,871,693

* Based on one month (2018/06) data

Data Preparation

- Is there really something worth exploring?
  - CFPM: Account accessed by User ID - xyz on 07/05/2018 02:45:19, no changes made.
  - DADD: Discount Plan Added, Eff date:07/05/2018, Exp date:No Expiration, Discount:MEPYDIS - Auto Pay Discount, User:abc

Data Preparation

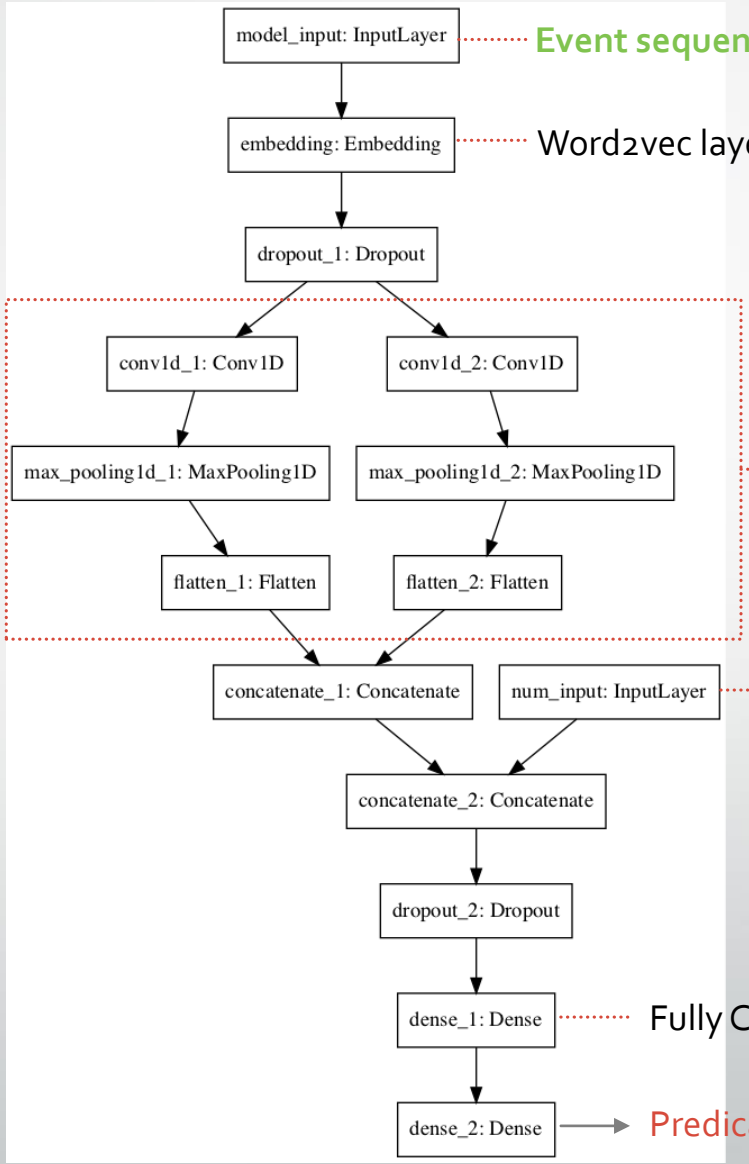
- Solution: concat a customer's Memo code into a string in the sequence of event time

	accountKey	subscriberKey	memo_events
0	110371590	34485218	VRFY,VRFY
1	110630572	3178571867274586306	RLPR,RLPR,RLPR,RLPR
2	110941250	55334237	CSMM,SEC1,RLPR,CSMM,RLPR,DADD,0007,DREM,SCNM,R...
3	110958555	1109585559139802363	RLPR,1024,1023,1024,1023,RLPR,1024,1023,SRMI,1...
4	111190833	22403572	0200,1023,VRFY

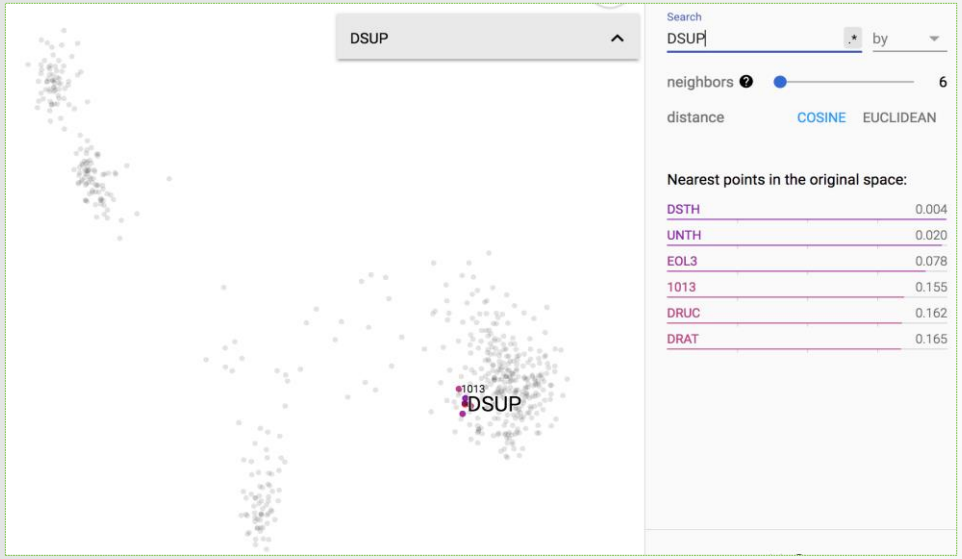
Data Preparation

- Target definition: customers who have “Deactivation” (churn) happens during Day 15 ~ Day 45 after month end.
- Target rate: 1.4%

Model Structure

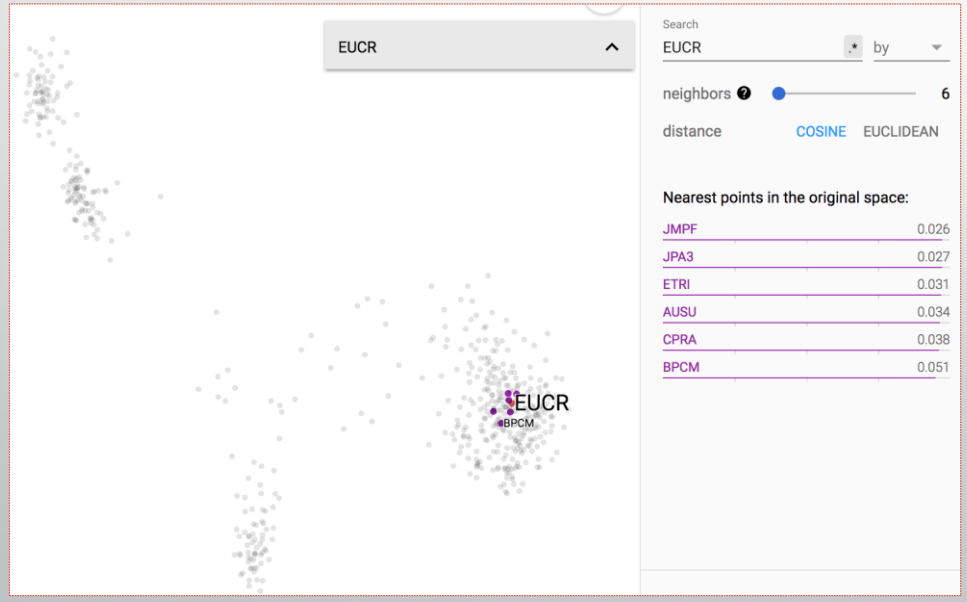


Word2vec based Event Embedding and t-SNE Visualization



Most similar events for 'DSUP':

DSUP	Data Up-sell message sent
DSTH	Data Throttle
UNTH	Data Un-Throttle
EOL3	SMS was sent to customer
1013	Reactivate Sub. feature
DRUC	Reached 100% Dom roaming
DRAT	XX% of roaming has been used

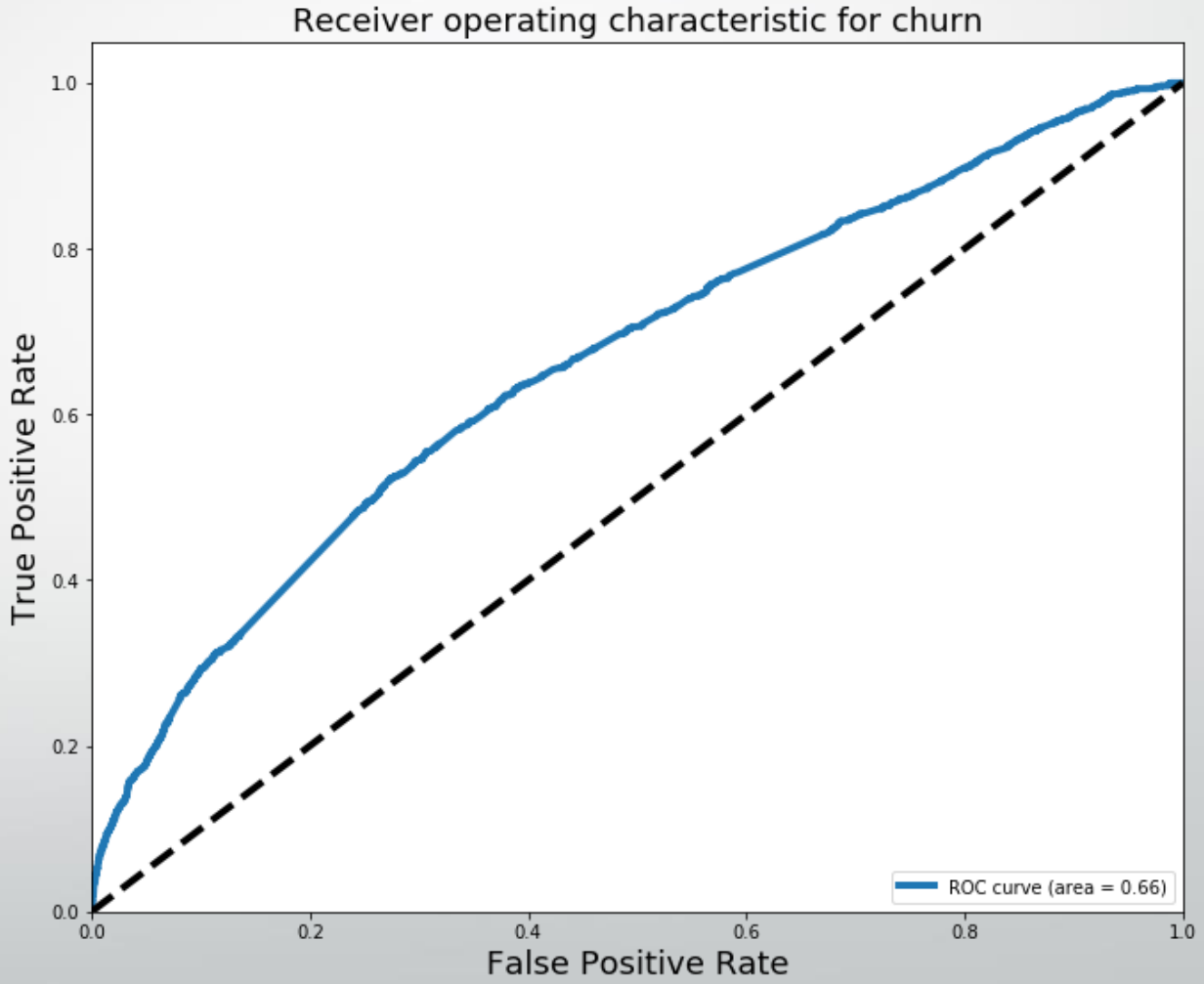


Most similar events for 'EUCR':

EUCR	Customer performed handset upgrade
JMPF	Jump upgrade
JPA3	iPhone upgrade SOC added
ETRI	Device Trade in Complete
AUSU	Jump tenure reset
CPRA	-
BPCM	Offer response (offer: XX, disposition: XX, reason: xx)

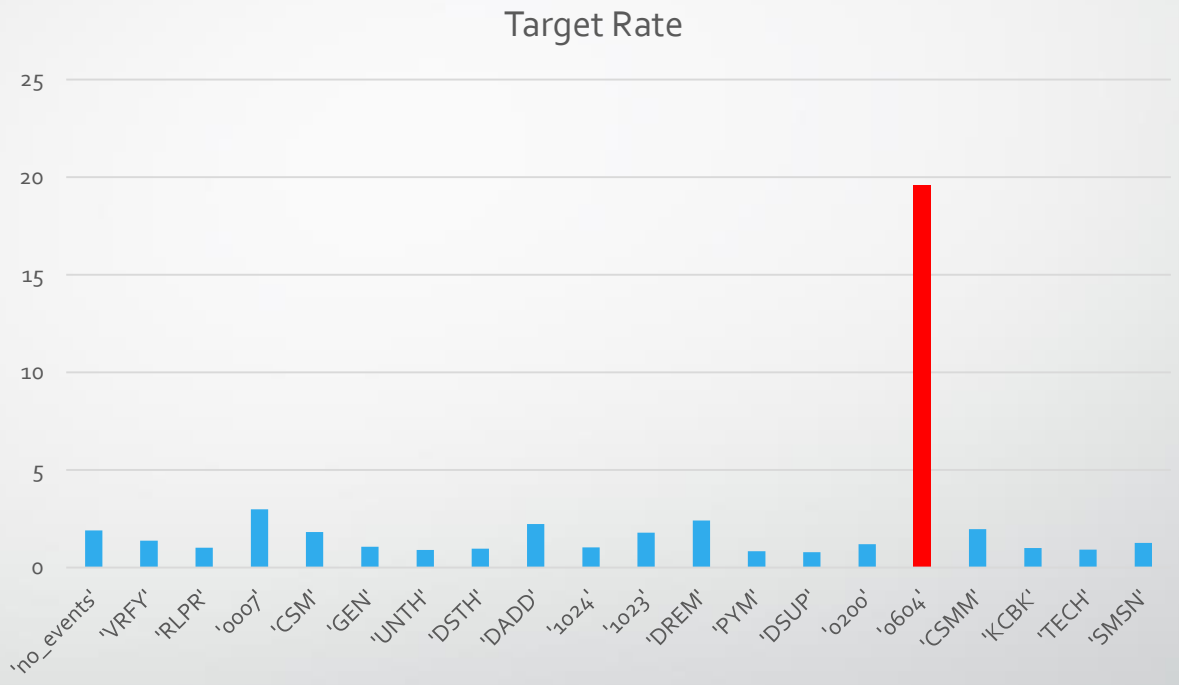
Model approach

First try:



Model Approach

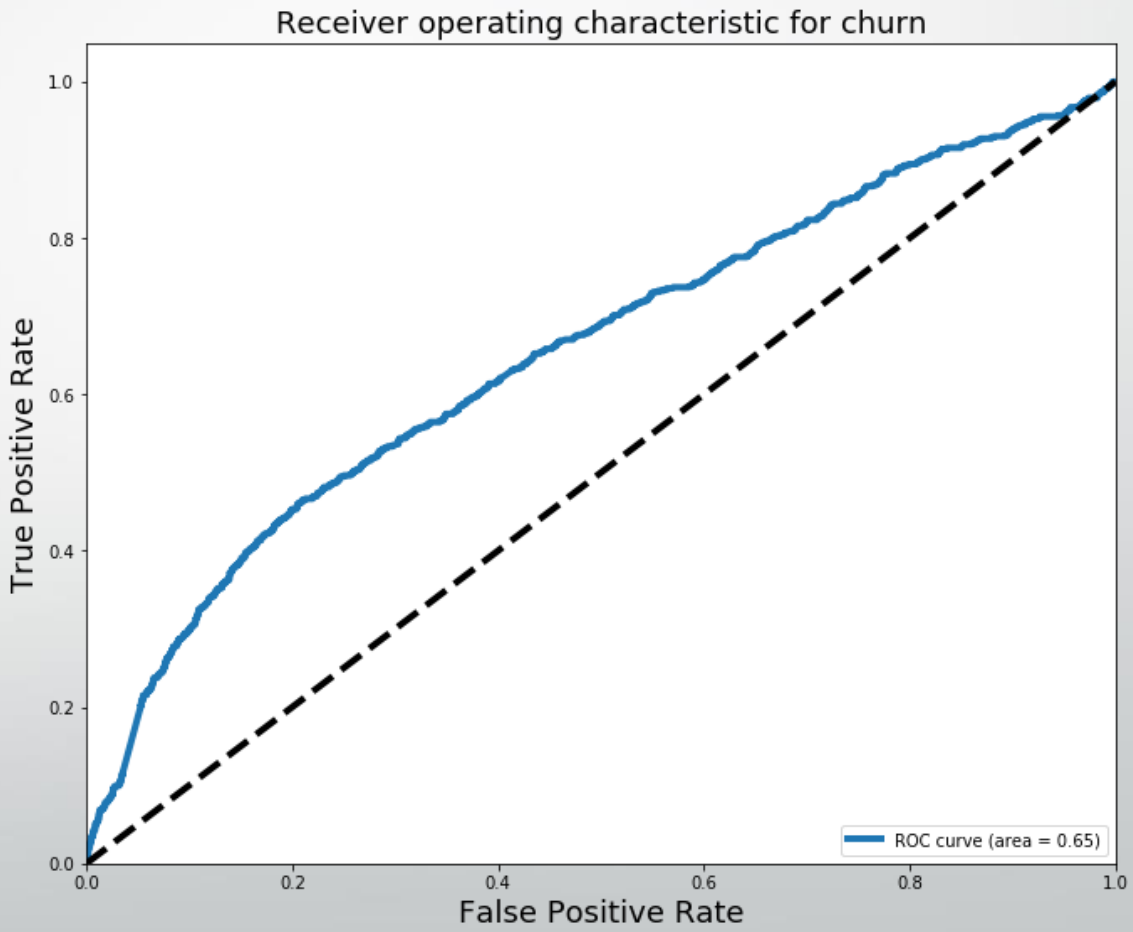
- Concern: is there potential target leakage?



0604: Future request. Activity: Cancellation.

Model approach

After removing 0604:



Model Approach

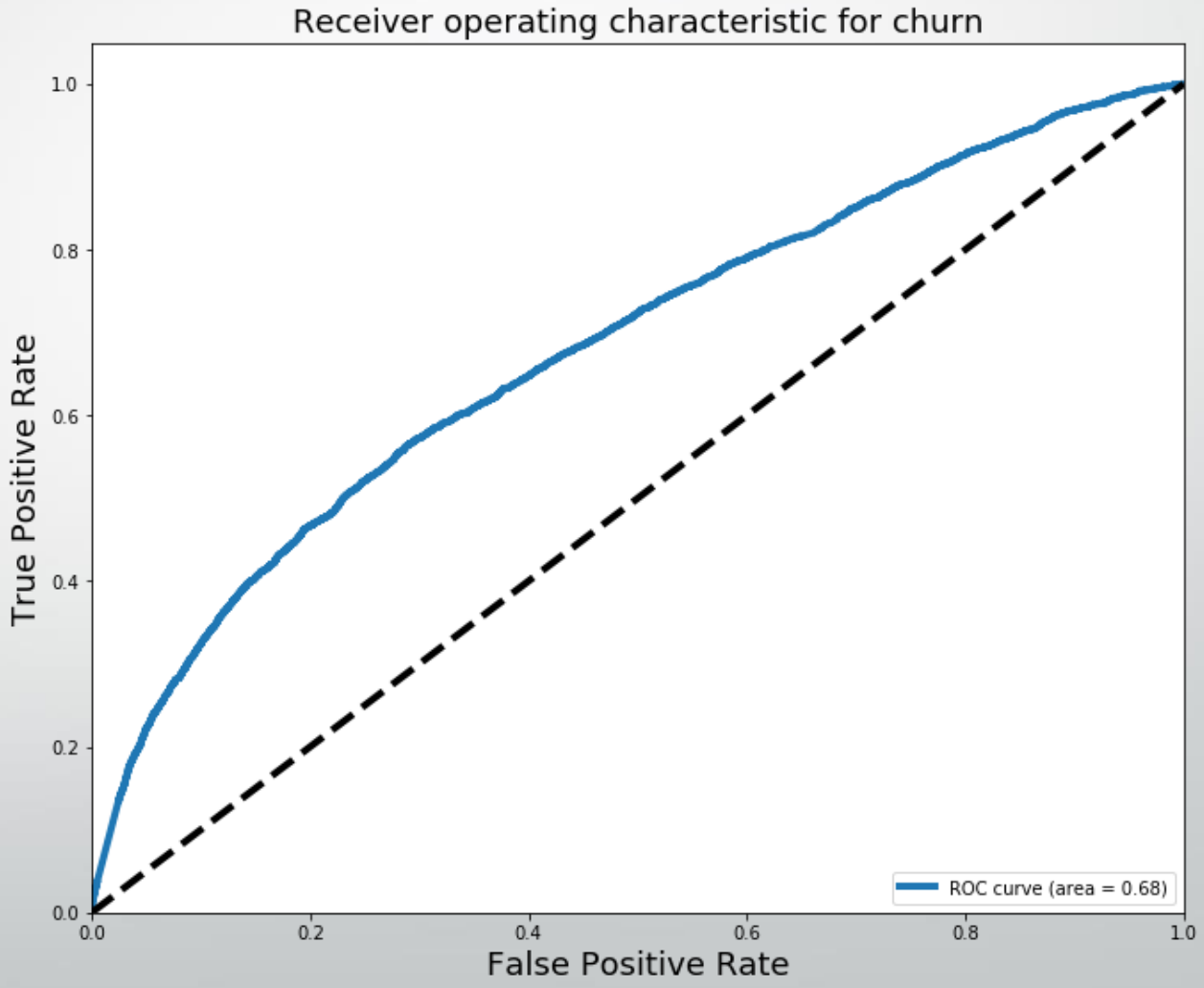
- What can we do to improve it?
- **Data**:
  - How many months do we need to include?
  - Down sampling?
- **Model**:
  - Word2vec embedding or not?
  - Padding length?

Model Approach

- What can we do to improve it?
- **Data**:
  - How many months do we need to include? – 4 months
  - Down sampling? – 1:2 (most effective)
- **Model**:
  - Word2vec embedding or not? - no Word2vec
  - Padding length? - 70

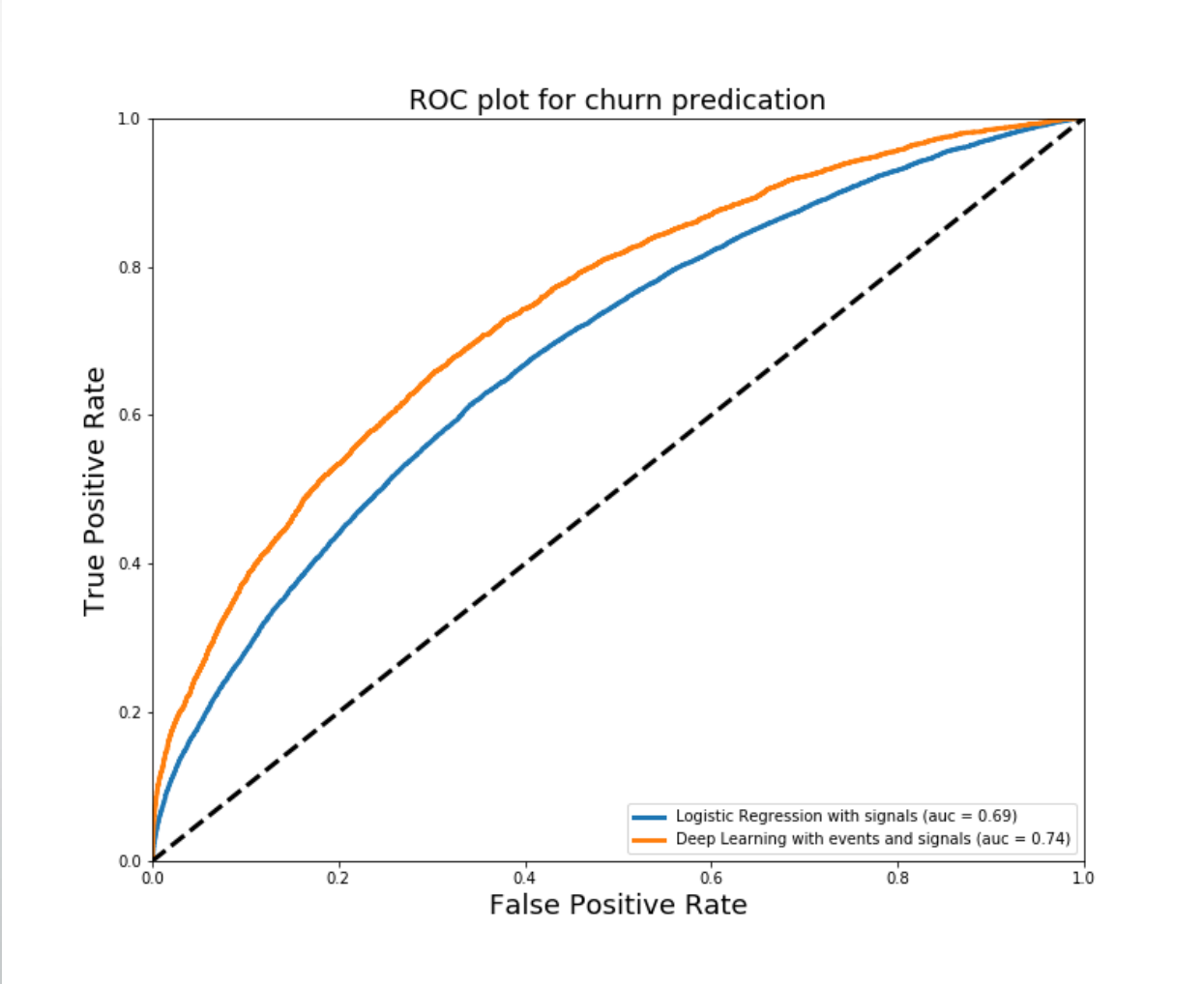
Model approach

Best result with text data only:



Model approach

Let’s add numerical features:



Summary

- What we learn:
- We developed an effective event sequence + CNN framework to predict churn.
- Data manipulation improves performance significantly, model tuning improves performance marginally.
- Next steps:
- Campaign sequence?
- Apply on different targets

Thank You!

- Questions?