

Big Contest NCSOFT

ASK-U

김진영 이태정 진명훈







기본적인 게임 이해

- ✓ 1998년 9월 1일부터 NCSOFT에서 서비스 중인 MMORPG
- ✓ 2016년 기준 누적 매출 3조 3천억, 전 세계 누적 이용자수 2천만명





문제 정의

이탈 유저를 예측하고 유저 일평균 결제금액에 비례하는 인센티브를 부여해서 잔존 상태로 유지시킬 수 있는가? n일 후에 이탈합니다!! 잔존 전환! 인센티브 부여

Q) 이탈 유저를 예측하고 유저 일평균 결제금액에 비례하는 인센티브를 부여해서 잔존 상태로 유지시킬 수 있는가?

우리는 위의 문제를 풀기 위해 아래 두 항목을 예측해야 합니다.

- ① 몇 일 후에 이탈할 것인지 or 이탈하지 않을 유저인지
- ② 유저가 관찰 기간 동안 일평균 얼마의 금액을 결제할지

그리고 우리는 위 문제를 풀기 전에 아래 두 명제를 정의해야 합니다.

- ① 이탈의 기준은?
- ② 유저들을 얼마나 관찰할 것인지?

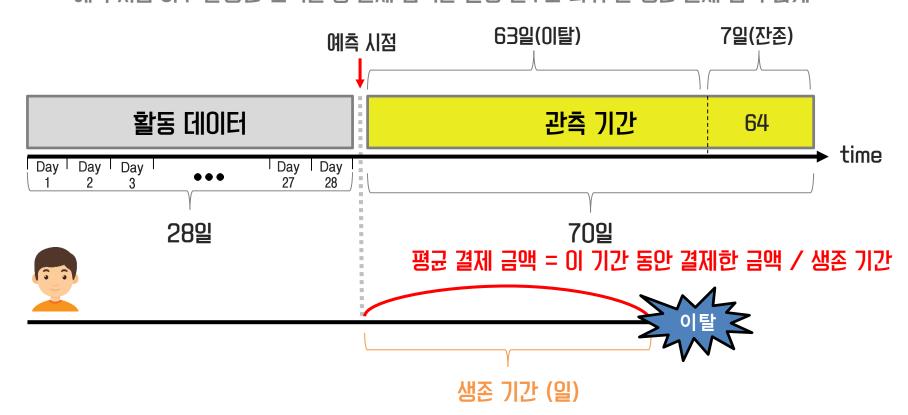


출처: NCSOFT

예측 시점에서 과거 28일간의 활동 데이터를 이용해서 모델 학습

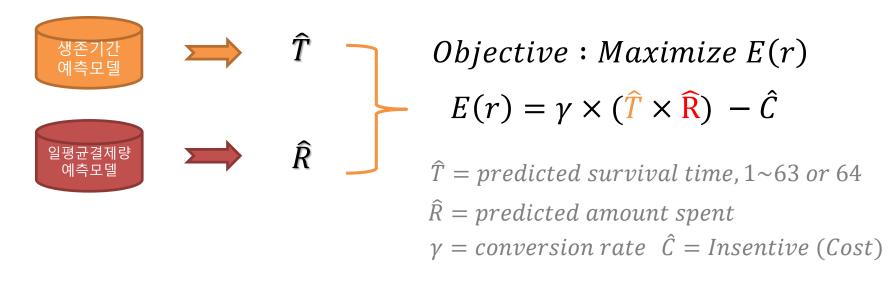
예측 시점 이후 70일 간의 관측을 통해 집계된 실제 고객별 이탈 시점(생존 기간) 및 평균 결제 금액 예측

- ✓ 63일 동안 이탈하지 않은 유저는 잔존으로 처리 (이탈 여부 판단 기간 7일 감안)
- ✓ 예측 시점 이후 발생한 고객별 총 결제 금액을 활동 일수로 나눠 일 평균 결제 금액 집계



만든 모델의 평가는 어떻게 할 것인가?

단순히 이탈 예측, 일평균 결제 금액 예측력을 높이는 것만이 목적이 아닌 자사의 기대이익을 최대화하는 것을 최종적인 목표로 삼음







40,000 Users

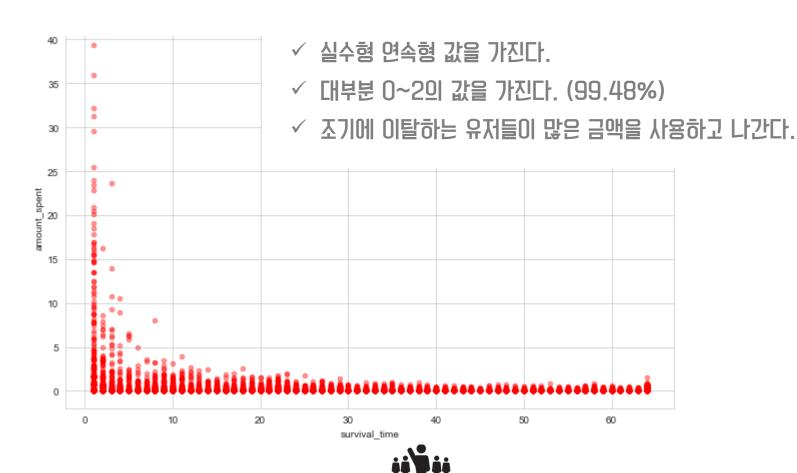
Non-Churn 54.99%

1. Survival Time: Multi Class

- ✓ 생존기간이 64일(잔존)인 유저의 비율이 55%
- ✓ 1~10일 그리고 29일에 이탈하는 유저가 많다.



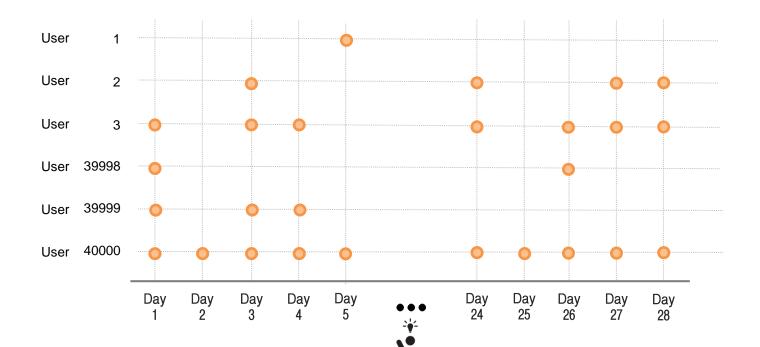
2. Amount Spent: Continuous Variable





3. Time dependent data

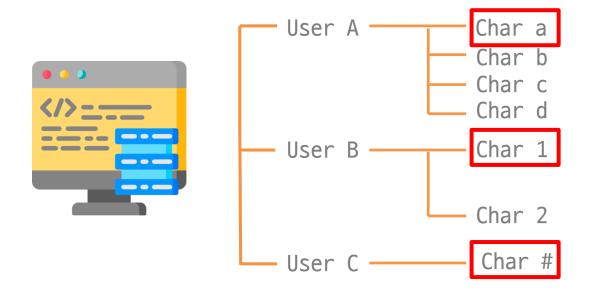
- ✓ 한 명의 유저가 최대 28일치의 관측치를 가진다.
- ✓ 매일 접속하는 유저가 있는 반면, 28일 중 이틀만 접속하는 유저도 존재한다.





4. User's Character

✓ 각 유저마다 최소 1개의 Character 활동 이력이 존재한다.





Content 2 데이터 전체리



저희는 다양한 가설을 설정해 성능을 향상시키려 시도했고 이 중 실제로 큰 향상을 보인 가설은 다음과 같습니다.

- 1. 주차별 Flatten
- 2. Count 변수 추가

목적: 시간에 강건한 모델 설계

3. 특정 예측 값에 가중치를 두어 학습 (Modeling Part에서 설명)

이 외 Feature로 추가된 다양한 가설들:

- 1. 죽고나서 부활을 하지 않는 유저는 게임을 포기할 가능성이 높다.
- 2. Login, Playtime이 적어지는 유저는 이탈 가능성이 높다.
- 3. 거래량이 많아지거나 Adena(게임 머니) 변화량이 증가하는 유저는 이탈 가능성이 높다
- 4. 익월 결제 금맥이 많을 유저는 마지막 주 활동에서 상대방에게 많이 공격받거니 혈맹(길드)의 활동이 증가한 유저일 가능성이 높다.





가설 1. 일자별 보다 주차별 데이터 Flatten이 더 효과적

- ✓ 시간에 강건한 모델을 구축하기 위해 일자별 데이터를 flatten시켜 주기로 결정
- ✓ 그러나 28일동안 거의 접속하지 않는 유저가 많음 > Data를 Sparse하게 만듦
- ✓ 때문에 사람의 생활 주기를 고려, 주차별로 Feature를 Flatten시켜 시간성을 고려

acc_id	week	day	Χ	Υ	Z
	_	1			
_	ı	2			
а					
	4	28			
l-	1	1			
b		5			



X_1		X_28	Y_1		Y_28	Z_1	•••	Z_28
Sp	arse Da	ata	Sp	arse Da	ata	Sp	arse Da	ata

X_1	X_2	X_3	X_4	Y_1	Y_2	Y_3	Y_4	Z_1	Z_2	Z_3	Z_4
	Data			Da	ata			Da	ata		





가설 2. Count 변수 추가

- ✓ 비모수적 접근방식
- ✓ 유저의 캐릭터별로 접속한 날이 같아도 활동한 플레이는 다름
- ✓ 일자별로 어떤 플레이를 했는지 count해서어떤 플레이 위주로 활동을 하는지 대표할 수 있는 변수 생성

acc_id	day	장사활동	전투활동
	1	0.0011	0.0023
	2	NaN or 0	0.4591
	3	0.0320	0.3336
	5	0.0220	NaN or 0
	7	0.1400	NaN or 0
a	9	1.2780	0.4802
	16	NaN or 0	0.6911
	21	0.9000	NaN or 0
	26	0.5000	NaN or 0
	28	0.2760	NaN or 0

acc_id	Avg(장사)	Avg(전투)	Count(장사)	Count(전투)
a	0.3936	0.3932	8 / 10	5 / 10





어떻게 Feature를 선정했는가?

- ✓ 제공받은 데이터 파일 5개를 acc_id 기준으로 결합
- ✓ 위의 Critical한 두 가설 및 기타 가설들을 기반으로 다양하게 feature 생성
- ✓ 특정 feature는 Char_id에서 주캐/부캐로 나누어 feature 추출 (주캐의 기준: 최고로 레벨이 높은 캐릭터)



Activity

Count

Playtime • 유저 별 일일 플레이 시간의 합 군집화

Login • 28일 간 접속 패턴 군집화 및 주차 별 접속 횟수

Char
 유저 별 캐릭터 수, 일 별 사용 캐릭터 수

Give up • 죽은 횟수 대비 부활 받은 비율 (부활시켜줄 친구가 있는지에 대한 사회 관계 지표)

Exp • Solo, Party, Quest에서 얻은 Exp로 군집화 후 numbering

Game
Money • 유저 단위, 캐릭터 단위로 game money 증/감 count
Change

Combat

Combat_std · 전투 횟수 및 유저 별 전투 데이터 편차(STD)

evel • 보유한 캐릭터 평균 레벨 및 성장 정도(max - min)

Class • 유저 별 보유한 클래스(직업)이 수

Pledge

Rank · 가입한 혈맹의 랭크

User • 유저 별 가입한 혈맹의 수 및 활동일

Server • a, b 서버 접속 횟수 👍

Trade

Item

• 판매 및 구매한 횟수, 아이템 수량, 금액

Trade type

• 거래 타입별(교환 창/개인 상점) count

User /Char

• 주력, 부수 캐릭터로 분리하여 Item, Trade type 생성

Trade time

• 일 평균 거래 횟수

• 주 거래 시간 및 playtime 대비 거래에 투자한 시간의 비율

User to other

• 판매/구매 상대의 수/합/비율 (장사캐릭터 간접적 판별)



Payment

Amount · 일 평균 결제량

 Pattern
 • 28일 간의 결제 패턴 군집화

• 28일 간 결제한 일 수





가설 검증을 위한 초기 모델 비교

• 전처리 단계에서 시간 강건성을 위해 세운 2개의 가설의 효과를 검증하기 위해 2개의 모델을 비교함

가설 미적용 모델

- 선택된 모델: XGB Regressor
- 선택 이유: DNN, LSTM, Random Forest 등의
 모델들과 비교했을 때 가장 높은 성능이 나옴

변수 형태

• 일 단위로 flatten한 데이터

VS

리더보드 결과

가설 적용한 모델

- · 선택된 모델: XGB Regressor
- 선택 이유: 동일

변수 형태

- 주 단위로 flatten한 데이터
- Count 변수 추가

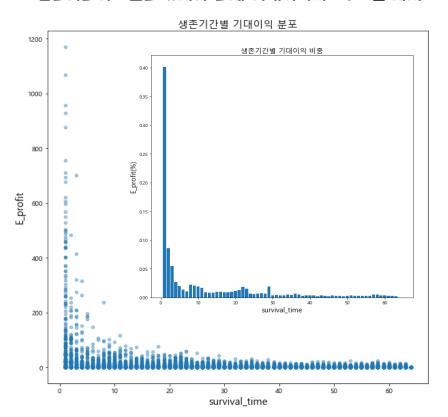
결과

기대이익(점)	X_test	test1	test2	Test1,2 합
가설 미적용 모델	4271	323	267	590
가설 적용 모델	2620 314	3346	2619	5966

- 가설 적용 모델이 X_test에서는 점수가 낮았으나, 리더보드에서는 10배 높은 점수를 얻음
- 주 단위 변수와 Count 변수가 시간에 강건한 변수임을 확인
- · 그럼에도 상위권(16000점 대)에 비해 낮은 점수이므로 성능 향상을 위한 `추가적인 전략'이 필요

문제 특성에 맞는 모델 설계: 예측 정확도 ≠기대이익

• 샘존기간이 1일인 유저가 전체 기대이익의 40%를 차지



• 생존기간이 실제 64이거나 64로 예측되는 경우 기대이익은 0

(부가 설명: 생존기간 64는 잔존으로 정의되어 인센티브에 따른 기대이익이 없다.)

Objective: Maximize
$$E(r)$$

$$E(r) = (T \times \hat{R}) \times \gamma - \hat{C} = \mathbf{0}$$
where $T = \begin{cases} 0, & \text{if } \hat{t} = 64 \text{ or } t = 64 \\ 30 \times e^{-\frac{(t-\hat{t})^2}{2 \times 15^2}}, & \text{otherwise} \end{cases}$

$$\hat{C} = 0.3 \times \hat{R} & \text{if } \hat{t} < 64 \text{ else } 0 = \mathbf{0}$$

$$\gamma = conversion \, rate$$

$$T = extended \, survival \, time$$

$$\hat{R} = predicted \, amount \, spent$$

$$\hat{t} = predicted \, survival \, time$$

기대이익을 최대화 하기 위한 전략

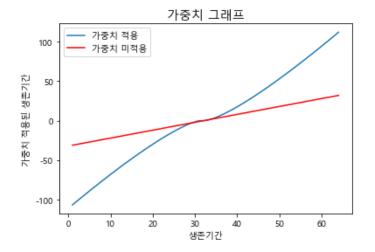
- 1. (생존기간 모델) 기대이익이 높은 유저가 몰려있는 생존기간 1일과, 기대이익이 0인 64일에 대한 예측 성능을 최대화 한다.
- 2. (일평균 결제량 모델) 기대이익이 높은 유저는 일평균 결제량도 높으므로, 높은 일평균 결제량 값의 예측 성능을 최대화 한다.
- 3. (가중치 적용) 특정 타켓 값 예측에 유리하도록 타켓에 가중치를 적용해서 1과 2의 전략을 현실화한다.

생존기간 예측 모델

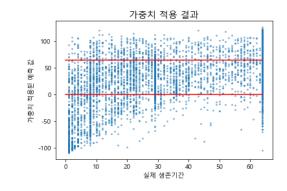
• 가중치 적용

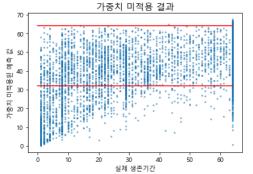
음의 가중치 : $(y-32) \cdot \ln(32-y)$, if y < 32 양의 가중치 : $(y-32) \cdot \ln(y-31)$, if $y \ge 32$

- 64일의 중간인 32일을 기준으로, 음과 양의 가중치를 적용
- 32와의 차가 클수록 더 큰 가중치를 적용하여, 1과 64가 가장 극적인 값을 갖도록 변환



- 가중치 적용 결과(X_test 기준)
 - (후처리) 예측 값이 1 이하인 경우 1일로, 64일 초과인 경우 64일로 변경
 - (성능 평가) 오분류된 값이 기대이익을 낮추므로 TP, TN을 모두 고려하는 정확도를 기준으로 평가
 - ✓ 생존기간 64일 (잔존) 예측 성능
 - 정확도 50.9%에서 78.3%로 27.4% 향상
 - ✓ 생존기간 1일 예측 성능
 - 정확도 83.5%에서 84.8%로 1.3% 향상
 - ✓ 기대이익(만점: 19262점)
 - 11370점에서 11671점으로 301점 향상





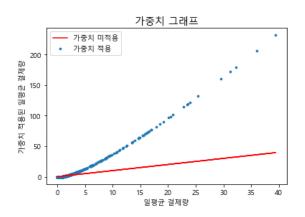


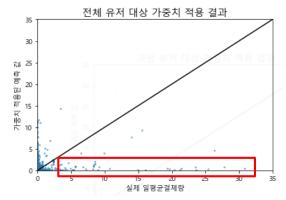
일평균 결제량 예측 모델

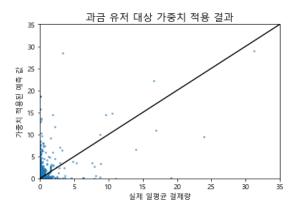
• 가중치 적용

가중치: 1.6·y·lny

- 자연로그를 통해 값이 클수록 더 큰 가중치가 적용되도록 함
- 그래프 모양과 rmse의 변화를 바탕으로, 휴리스틱적인 방법으로 1.6이라는 상수 곱을 결정
- 추가 전략: 무과금 유저를 제외
 - 무과금 유저가 전체 유저 40000명의 40%를 차지함.
 - 이로 인해 가중치를 적용해도 모든 예측 값이 이에 가까워지는 현상이 발생하여 가금 유저만을 대상으로 모델 학습
- 가중치 및 추가 전략 적용 결과(X_test 기준)
 - (후처리) 예측 값이 0 미만인 경우 0으로 변경
 - ✓ 1 이상의 고맥 일평균 결제량(상위 1%)의 rmse
 - rmse 5.717에서 4.940으로 0.777 개선
 - ✓ 기대이익(만점: 19262점)
 - 5807점에서 7274점으로 1467점 향상







Content 4 결 론

종합 결과(리더보드 제출 결과)

- · 가중치를 적용한 결과, 리더보드 상의 기대이익 총합이 5966점에서 13400점으로 상승
 - Q) 예측에 영향을 미치는 주요 특성은 무엇인가?

	공통	생존기간	일평균 결제량	
접속	Playtime_4	Playtime_count	Playtime_3	
留书	Plagtille_4	Plagtille_count	Day_sum	
		Solo_exp_count		
플레이	_	Quest_exp_count	Playtime	
글데이	_	Fishing_count	Playtime	
		Level		
	Plus_am			
게임머니	Minus_am	Amount_diff	Game_money_change_count	
	Game_money_change_4			
결제	_	_	Pay_mean	
는 All	-	-	Amount_spent_1	

- 1. 생존기간 및 일평균 결제량 예측
 - 모두 넷째 주의 플레이 시간과 아데나 변동량이 유의 (특정 패턴을 보임)
- 2. 생존기간 예측
 - · count 변수들이 주요 속성으로 작용
 - · 접속빈도가 높지만 다양한 컨텐츠를 즐기지 못하고 대부분 solo 사냥을 통한 Level up에만 포커싱
- 3. 일평균 결제량 예측
 - · 첫째 주에 결제량이 주요 속성으로 작용
- 4. 게임머니 변동량
 - 이탈 유저의 경우, 아이템 거래량이 상당히 낮으며, 4주차로 갈수록 일일 아데나 변동량이 점점 낮아진다
 - 결제량이 경우, 아데나 변동량이 적을수록 결제량이 높은 수치 기록





감사합니다