

우리는 대회에서 제시된 정규화 된 데이터를 보고 먼저 복원을 시도했다. 정규화 되어있기 때문에 기존의 특징들을 알기가 어려웠고 그에 따라 데이터를 통해 세계를 읽는 것 또한 어렵기 때문이다. 아이템의 경우에는 최소 변화 단위가 1이기 때문에 가장 작은 값을 0, 그 다음 값을 1로 보았으며, 경험치의 경우 블레이드&소울을 플레이했던 경험과 파워북의 도움으로 최소 획득 경험치량은 6이며, 최소 퀘스트 경험치는 70임을 토대로 복원을 했다. 그 후 Party데이터의 경우 각 acc\_id별로 각 주 별로 파티를 몇 분, 몇 회, 몇 명과 하였는지를 구했고, Guild데이터의 경우 각 acc\_id가 문파에 속해 있는지, 이중문파인지, 문파의 규모가 큰지 작은지를 구분하였다. 마지막으로 Trade데이터는 각 주 별로 누적 건넨 재화량을 구하여 모두 activity와 합하여 최초 전처리를 마무리했다. 우리는 여러가지 방식으로 접근하기 위해서 활동주로 나눈 주당 평균 데이터와 각 column을 주 별로 나타낸 column을 펼친 데이터 2가지를 만들었다.

전처리를 끝낸 데이터를 찬찬히 읽어보니 거대한 쌍봉분포를 띄고 있다는 것을 파악할 수 있었다. 이 쌍봉은 각각의 큰 특징이 있는데 하나는 사냥을 지나치게 많이 하지만 숙련 던전과 레이드는 절대 돌지 않았고, 다른 하나는 다양한 콘텐츠를 자유롭게 즐기고 있었다. 그를 통해 이들이 게임 속 매크로계정인 Auto계정과 사람계정임을 알 수 있었다. 때문에 우리는 데이터를 군집화하기로 결정하였다.

그 군집은 크게 2가지로 Auto계정, 사람 계정이며 하부 군집은 다음과 같다. Auto계정은 아이템 등 재화를 사냥, 인던을 통해 수집하는 개미 계층, 그 재화를 모으는 관리자 계층, 관리자로부터 잡화나 금화를 받아 일반인에게 현금을 받고 판매하는 상인 계층이 있다. 사람계정은 그저 즐겁게 게임을 즐기는 일반인 계층, 게임에 할애하는 시간이 아주 큰 하드코어 계층, 더 나아가 승객으로부터 요금을 받고 레이드 및 숙련 던전을 클리어해주는 버스기사 계층이 있으며, 각 계층 안에 고속성장캐릭터가 존재함을 파악했고, Auto계정과 사람계정의 예외로 튜토리얼만 완료한 더미계정이 숨어 있었다.

이들을 분류하기 위해 먼저 전체 데이터로부터 Auto계정 조건을 추려 뽑아내기로 했다. 게임 플레이 경험상 Auto계정들을 아주 많이 보았으며, 이들이 출몰하는 지역의 특징이나 행동 패턴으로부터 이들을 추려 낼 수 있는데 기본적으로 Auto계정은 재화수집이 주 목적이다. 물자이동을 위해 거래가 반드시 발생하게 되기 때문에 먼저 Trade데이터로부터 Auto 계정을 추려낸다.

일반적인 거래는 다음과 같다. Source 계정이 물건을 건내면 Target 계정은 금화를 지급한다. 하지만 Auto계정은 Source 계정이 물건도 건내고 금화도 건내며 Target 계정은 그저 받기만 한다. 이에 근거하여 반복적으로 물건과 금화를 동시에 건내는 계정목록과 받는 계정목록을 추려낸다. 이 특이성은 우리가 최초에 Auto계정이 있다는 사실을 파악하게 된 근본적인 원인이기도 하다.

따라서 Auto 계정의 가장 중요한 정보는 '반복적으로 소수의 target에 잡화, 금화를 공급했는가?'이다.

우리가 원하는 것은 전체 Auto 계정의 목록이기 때문에 먼저 Train과 Test 데이터를 합쳐 Total Trade 데이터를 만들었다. Total 데이터에 포함된 거래정보는 총 6,848,159건인데 이중 2,871,734

건이 Source계정이 물건도 건내고 금화도 건낸 거래정보다. 여기서 Source계정의 수는 106,926개이며 Target계정의 수는 9,897개다. 이 목록이 가장 초기 Auto계정 의심 목록이며 다만 Target계정이 Source가 될 수도 있으므로 Source와 Target의 교집합이 존재하며 이 교집합은 재미이면서 관리자일수가 있다. 따라서 교집합을 Source에서 제거해줘야 Source에 최하층의 재미의 비율이 최대화된단다. 다만 이 경우에도 Source 목록에 단일 품목으로만 재화를 모으는 계정이 있을 수 있으며 실제로 확인했을 때 존재했다.

('a7dca722d8908a1a76b0d93b283b483fbe6932c7f6ad9b7d2139942ea3c1ca5a')

이 때 Source는 101,491개, Target은 9,897개다. 또 Auto계정은 반드시 이 목록 안에 있지만, 사람계정이 지인을 돕는 차원에서 재화를 지급하는 경우도 있기에 Activity 데이터로부터 한번 더 추려내기로 했다.

Activity 데이터는 총 14만건이 있는데 위의 101,491개의 Source계정과 대조결과 32,710개의 계정에 대한 정보를 갖고 있으며, Target계정과 대조결과 6,292개의 계정 정보를 갖고 있었다.

Auto 계정의 경우 사냥 매크로이거나 관리자, 상인이기 때문에 할 수 없거나 하지 않는 콘텐츠들이 있다. 그것은 Raid, RaidLight, Skilled\_inzone, 그리고 문파, 생산, payment이다. 또 Source는 봉인부적을 구매해서 장비를 옮기지는 않으므로 잡화, 금화의 거래량은 0이면 안되며 그 외 물품 거래량은 0이어야 한다. 실제로 잡화, 금화 외 다른 품목이 0이 아닌 acc\_id를 직접 확인했을 때 auto계정의 특징을 갖고 있지 않았다.

위의 조건을 적용하면 Source는 30,421개(전체 21.73%), Target은 1803개(전체 1.29%)가 남는다.

Source는 사냥매크로이므로 Solo inzone, Light inzone, Normal inzone 데이터가 모두 0일수는 없다. 3가지 특징 모두 주 평균 1회 이하이고(인던을 돌지 않고 퀘스트만 클리어하는 매크로는 수치가 낮음) 주 평균 전투시간이 작음(<30분) acc\_id를 제거한다.

또 매크로이므로 거래량이 지나치게 작음(금화 거래량<100) acc\_id도 제거한다. (auto 판별에 있어 가장 중요한 정보는 '반복적으로 소수의 target에 잡화, 금화를 공급했는가?' 이므로 기준을 낮게 잡아야만 한다. 또한 사냥만 반복하고 Auto 관리자 계층에게 재화를 아직 공급하지 않은 계정도 많으며 매크로라고 해서 항상 효율적으로 동작하지는 않는다. 게임 속에서 가만히 정지해 있는 매크로 캐릭터도 아주 많이 보이는 것이 그 증거다.)

trade_week	trade_day	trade_time	source_acc_id	target_acc_id	item_type	item_amount
1	7	01:24:37	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	grocery	39.0000
1	7	01:24:37	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	money	39.7780
2	1	00:09:12	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	grocery	1.0000
2	1	00:09:12	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	money	12.8538
2	4	04:44:17	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	grocery	1.0000
2	4	04:44:17	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	money	4.6871
2	4	20:56:37	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	grocery	1.0000
2	4	20:56:37	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	money	4.6839
2	6	01:04:44	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	grocery	3.0000
2	6	01:04:44	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	money	4.6816
2	6	21:38:34	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	grocery	1.0000
2	6	21:38:34	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	money	4.6812
3	1	00:37:09	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	grocery	1.0000
3	1	00:37:09	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	money	4.6822
3	3	00:43:35	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	grocery	1.0000
3	3	00:43:35	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	money	4.6830
3	3	21:59:00	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	grocery	1.0000
3	3	21:59:00	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	money	4.6854
3	6	04:05:23	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	grocery	1.0000
3	6	04:05:23	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	money	4.6853
3	7	00:50:03	646cf0f6a7ff9fe50854556f71c6abd694bd4d4b72df97...	3dd1b99940b0a4a01a72586441a2dcf6fb80e19879f08e...	grocery	2.0000

이 아이디는 주당 평균 solo, light, normal 클리어 수치가 3이하이지만 auto 계정의 근거인 거래 정보가 위와 같다.

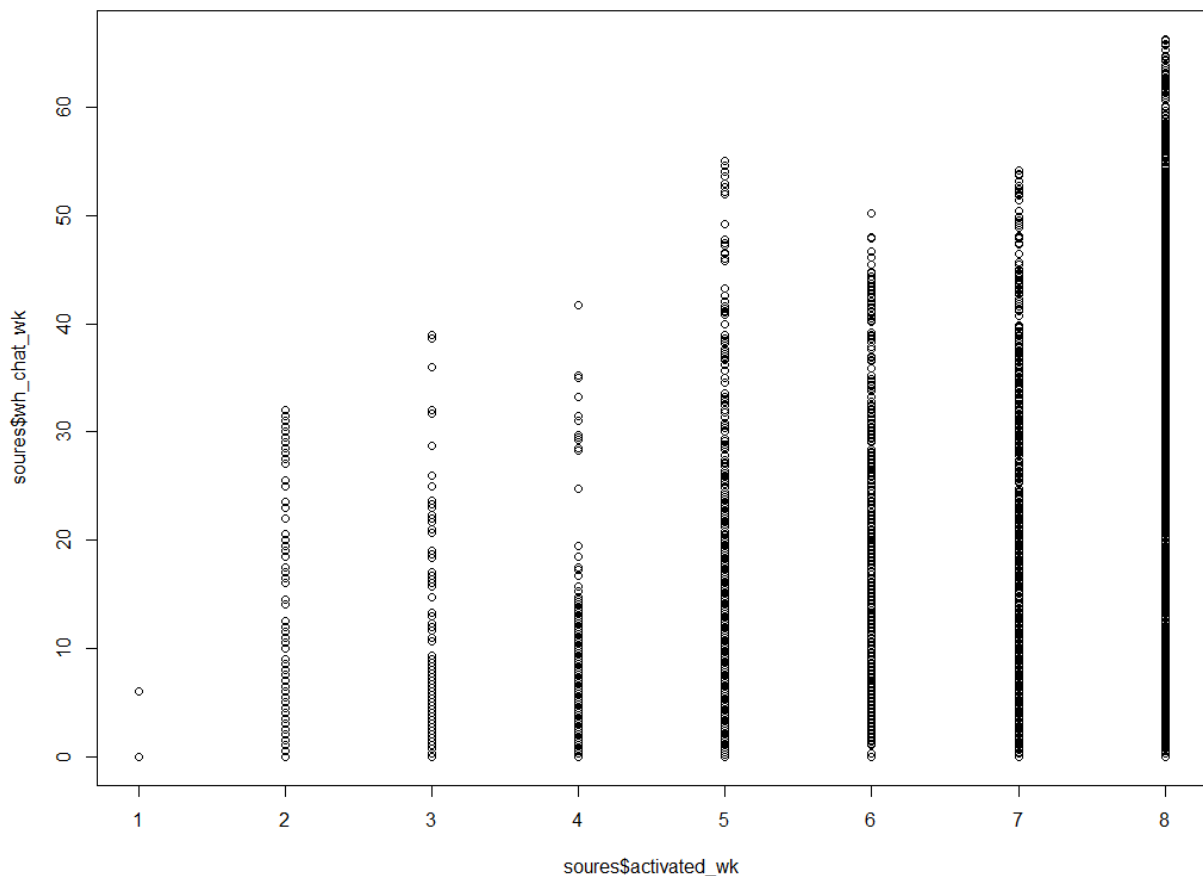
개략적으로 사람계정이라고 생각되는 acc\_id를 제거한 결과 Source는 26,659개(전체 19.04%)가 되고, Target은 1803개(전체 1.29%)가 된다.

이를 train test로 나누면 Train에는 Source 18957개(18.96%), Target 1278개(1.28%), Test에는 Source 7702개(19.25%) Target 525개(1.31%)가 있다.

여기까지 했을 때 재밌는 점이 있는데 이 매크로들의 행동 트리거가 귓속말이 아닐까 라는 가설을 세울 수 있었다.

Filter												
he_wk	money_wk	nl_chat_wk	wh_chat_wk	dis_chat_wk	guild_chat_wk	buff_cnt_wk	making_wk	solo	light	skilled	norma	
3	2527.23	15.5718	0.00	49.29	0	0	572.86	0	519.29	199.57	0	1'
3	2609.81	15.0306	0.00	45.00	0	0	745.14	0	503.36	187.71	0	1h
3	2947.40	14.8323	0.00	44.88	0	0	837.25	0	463.38	187.12	0	1h
3	2576.29	15.2602	0.00	47.12	0	0	701.12	0	526.06	186.94	0	1h
7	2475.91	14.7007	0.00	48.62	0	0	859.12	0	461.06	179.00	0	1'
7	2580.45	14.5525	0.00	48.50	0	0	687.62	0	458.69	178.75	0	1'
7	2440.24	14.3517	0.00	48.75	0	0	633.12	0	458.62	177.69	0	1'
7	3026.82	14.1327	0.00	40.25	0	0	789.62	0	448.69	175.94	0	1'
7	2943.32	13.9093	0.00	40.00	0	0	899.62	0	439.38	174.00	0	1'
7	2793.69	14.0072	0.00	39.25	0	0	797.62	0	469.00	173.81	0	1'
7	2803.90	13.7956	0.00	39.88	0	0	675.12	0	491.19	173.12	0	1'
7	2194.88	13.2227	0.00	42.38	0	0	634.88	0	436.69	172.31	0	1'
7	2426.76	13.5773	0.00	45.12	0	0	652.00	0	426.69	171.81	0	1'
7	2427.44	13.5392	0.00	40.12	0	0	564.00	0	455.06	170.50	0	1'
5	2169.10	13.0310	0.00	40.25	0	0	624.38	0	423.06	169.44	0	1h
5	2464.14	14.3209	0.00	41.75	0	0	732.25	0	470.75	169.31	0	1h
5	2223.22	13.9425	0.00	42.25	0	0	534.62	0	447.56	169.06	0	1h
5	2295.21	13.4912	0.00	42.25	0	0	590.38	0	447.44	169.00	0	1h

행동이 비슷한 acc\_id는 깃속말 횟수도 비슷하다는 것과 아래 플롯이 그 근거다.



활동주가 길수록 깃속말 횟수도 증가하는 양상을 보인다.

마지막으로 target 아이디의 target중 auto가 아닌 아이디를 추리고 그 중에서 money만을 거래한 캐릭터들을 모으면 골드 구매자 목록이 된다. 이들은 현금으로 골드를 구매한 것이며 hard 유저로 볼 수 있어 hard 군집으로 분류하였다.

그리고 사람 계정의 경우에도 trade 데이터를 읽어보던 중 특이한 집단을 발견할 수 있었다. 하나의 source로부터 많은 양의 금화를 받고, 그 금화를 몇 분 후에 다른 target 다수와 동일하게 나눠가지는 집단을 발견한 것이다. 이들은 버스기사이다. 또, source가 여러명이고 이를 다수와 나눠가지는 경우는 레이드를 클리어하고 나서 전리품 경매를 통한 이익잉여금을 나눠가지는 것이다. 그러므로 이 source와 target들은 1회이상 본인의 능력으로 레이드를 클리어했다고 볼 수 있으며, 본인의 능력으로 레이드 파티의 일원이 될 수 있다면 게임에 능숙하다는 뜻이 된다. 따라서 위의 hard와 통합하여 hard&bus 군집으로 분류하였다.

이 hard&bus군집은 Total 10530개다.

예외로 앞서 언급했던 dummy 계정이 있는데 이 계정은 튜토리얼만 클리어 한 계정들이다. 즉, 대나무 마을에서 은광일지부장을 잡는 퀘스트까지만 클리어한 것으로 퀘스트 경험치 획득량이 8610이다.

이들은 선분류한 auto에도 있고 사람계정에도 있는데, 데이터 상 차이가 전혀 없기에 kmeans로 비지도 학습을 해도 1개의 클래스로 몰리며 어디로 분류해야 할 지 알 수 없다. 그래서 선분류한 목록에서 이 더미 계정들을 제거한다. 이들이 정확도를 올리는데 있어 가장 큰 노이즈가 된다.

따라서 태생적으로 일정 수치 이상 정확도를 올릴 수 없게 되어있다.

그 결과 Auto계정은 총 28,366개이며, 사람 계정은 87,610개, 더미계정은 24,024개이다.

또 다른 특이한 군집이 있는데 이들은 고속성장권을 사용한 계정으로 퀘스트 홍문 경험치 획득량이 4,500,000인 acc\_id들로 고속성장캐릭터가 튜토리얼 완료를 했을 때 받는 경험치량과 일치한다. 이들은 다양한 군집에 교집합을 가지며 총 16,410개가 있다.