6조 최종발표

이진규
전민재
홍동준
차정우

Index

Feature Engineering

- Clickstreams Data
  - Reference
  - Word2vec
  - Trans2vec
  - LDA
  - Clustering
  - Tanimoto Similarity
- Searchkeywords
  - Trans2vec
  - Cosine Similarity

Feature Engineering_ClickStreams

〈Table 2〉 Summary of User's Profile

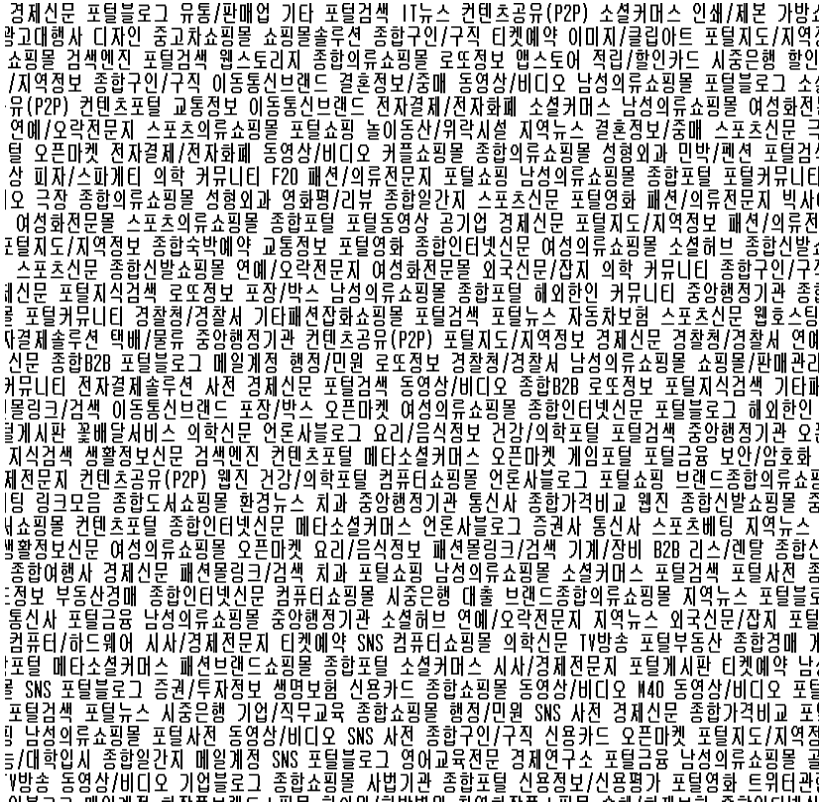
		Variable description and the number of variables		Author and year
Behavior	Preferences of websites	Ratio of pagevews for website category	22	De Bock and Van den, 2009 Goel et al, 2012
		Coefficient of variation for website category	1	
	Usage Pattern	Total number of pageviews for website category	1	De Bock and Van den, 2009 Eleonora, 2013 Goel et al., 2012
		The total number of days to visit	1	
		Ratio of pagevews for time	4	
		Ratio of pagevews for day	7	
		Ratio of pagevews for month	12	
		Coefficient of variation for time	1	
		Coefficient of variation for day	1	
		Coefficient of variation for month	1	
	Search Behavior	Total number of search keywords	1	Gallagher and Parsons, 1997 Jones et al., 2007 Murray and Durrell, 2000
	Interest	Ratio of pagevews for news website category	12	Yoonjin Hyun et al. ,2015
Demographics		Gender(2)	1	Baglioni et al., 2003 De Bock and Van den, 2009 Eleonora, 2013 Murray and Durrell, 2000 Jones et al., 2007
		Age(5)	1	De Bock and Van den, 2009 Eleonora, 2013 Murray and Durrell, 2000
		Marital Status(2)	1	Eleonora, 2013 Murray and Durrell, 2000
		Residence(13)	1	Eleonora, 2013
		Job(20)	1	De Bock and Van den, 2009 Eleonora, 2013

* The figures in parentheses are the number of categories for each variable.

사용자의 프로파일(profile) 생성

- SITE_CNT
- ST_TIME
- Coefficient of Variance

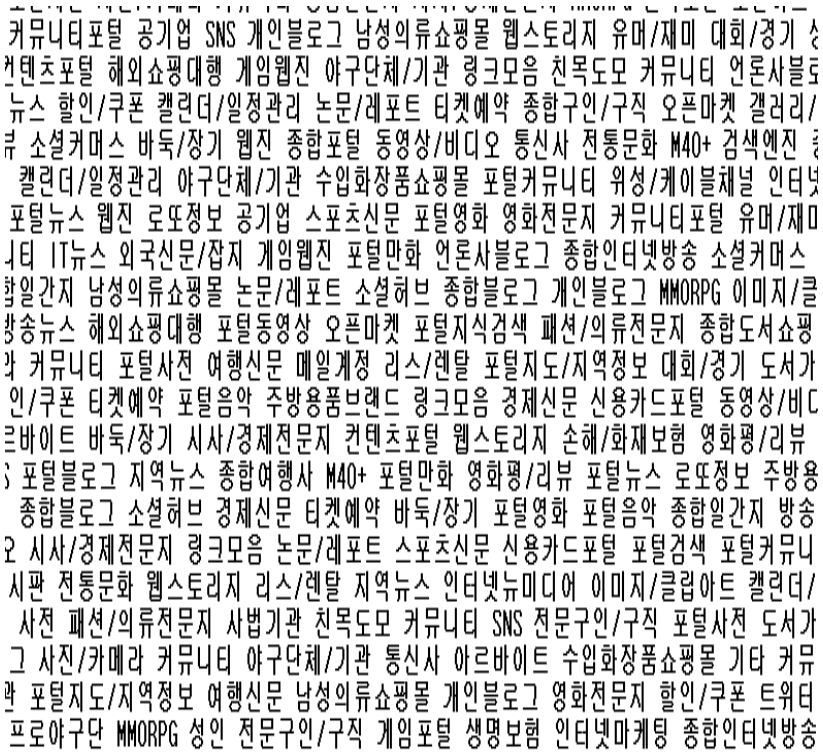
Feature Engineering_ClickStreams



word2vec

M40	M30	F30	F40	F20	M20
0.05402084	0.078478034	0.11684876	0.04621761	0.09891620	0.051973920
0.07036006	0.083267641	0.11603718	0.04122877	0.12021978	0.038371077
0.03383840	0.100421427	0.18343648	0.01470120	0.08068151	0.003380634
0.10046832	0.033595364	0.07295731	0.08625787	0.15970943	0.101459094
0.08735713	0.008872623	0.10778582	0.08988349	0.11100907	0.046284431
0.09337527	0.046527203	0.09669757	0.04247791	0.01120579	-0.053026004

Feature Engineering_ClickStreams



trans2vec

M40	M30	F30	F40	F20	M20
0.05402084	0.078478034	0.11684876	0.04621761	0.09891620	0.051973920
0.07036006	0.083267641	0.11603718	0.04122877	0.12021978	0.038371077
0.03383840	0.100421427	0.18343648	0.01470120	0.08068151	0.003380634
0.10046832	0.033595364	0.07295731	0.08625787	0.15970943	0.101459094
0.08735713	0.008872623	0.10778582	0.08988349	0.11100907	0.046284431
0.09337527	0.046527203	0.09669757	0.04247791	0.01120579	-0.053026004

Feature Engineering_ClickStreams

```
f <- function(x, dt) {
  itemfreq <- table(dt[CUS_ID==x, ACT_NM])
  fitems <- itemfreq[itemfreq >= 1]
  act <- names(fitems)
  return(paste(act, collapse = " "))
}
md.dt$ACT_NM <- gsub(" ", "_", md.dt$ACT_NM);
tr.t.dt$ACT_NM <- gsub(" ", "_", tr.t.dt$ACT_NM)
items <- unlist(sapply(cs.dt$CUS_ID, f, md.dt))
items <- c(items, unlist(sapply(unique(tr.t.dt$CUS_ID), f, tr.t.dt)))

##### create the document term matrix (DTM)
# DTM is a mathematical matrix that describes the frequency of terms that occur in a collection of documents.
# In a document-term matrix, rows correspond to documents in the collection and columns correspond to terms.

tic <- proc.time()
items.dtm <- DocumentTermMatrix(Corpus(VectorSource(items)))
print(proc.time() - tic)

##### Run LDA model

tic <- proc.time()
lda.model <- LDA(items.dtm, k=20, method="Gibbs", control=list(burnin=1000, iter=1000, keep=50))
save(lda.model,file="lda.model.rda")

print(proc.time() - tic)
# Saving and loading lda model:
# saveRDS(lda.model, "lda_model.rds")
# lda.model <- readRDS("lda_model.rds")


##### Calculate the per document probabilities of the topics

items.theta <- as.data.frame(posterior(lda.model)$topics)
#head(items.theta[1:5])
train <- cbind(data.frame(CUS_ID=cs.dt$CUS_ID), items.theta[1:2500,])

# for test data
test.CUS_ID <- unique(tr.t.dt$CUS_ID)
test <- cbind(data.frame(CUS_ID=test.CUS_ID), items.theta[2501:5000,])
```

LDA

ACT_NM

CUS_ID	ACT1	ACT2	ACT3	ACT4	ACT5	ACT6	ACT7	ACT8	ACT9	ACT10
1	1	0.03846154	0.03846154	0.05029586	0.07396450	0.01479290	0.014792899	0.05621302	0.03254438	0.02662722	0.06213018
2	2	0.01198630	0.02910959	0.05650685	0.04623288	0.03253425	0.008561644	0.08732877	0.01541096	0.05650685	0.03938356
3	3	0.03735632	0.04885057	0.01436782	0.03160920	0.04310345	0.014367816	0.11206897	0.09482759	0.07183908	0.01436782
4	4	0.02201258	0.02201258	0.02830189	0.04716981	0.04088050	0.015723270	0.13522013	0.12893082	0.02830189	0.04088050
5	5	0.05895197	0.02838428	0.04585153	0.03275109	0.01528384	0.010917031	0.03275109	0.06331878	0.18122271	0.02838428
6	6	0.06250000	0.07083333	0.07916667	0.07083333	0.02916667	0.020833333	0.03750000	0.06250000	0.03750000	0.11250000

	ACT11	ACT12	ACT13	ACT14	ACT15	ACT16	ACT17	ACT18	ACT19	ACT20
1	0.01479290	0.02071006	0.10946746	0.05621302	0.07988166	0.09171598	0.12130178	0.02071006	0.03254438	0.04437870
2	0.05993151	0.10102740	0.07705479	0.02226027	0.01541096	0.01198630	0.03253425	0.03595890	0.19349315	0.06678082
3	0.03735632	0.12356322	0.02586207	0.02586207	0.07183908	0.01436782	0.02011494	0.07183908	0.03735632	0.08908046
4	0.02830189	0.04088050	0.08490566	0.04088050	0.03459119	0.03459119	0.04088050	0.02201258	0.14150943	0.02201258
5	0.05458515	0.03711790	0.02838428	0.09388646	0.05021834	0.05458515	0.05021834	0.05895197	0.02838428	0.04585153
6	0.02916667	0.02916667	0.04583333	0.07083333	0.08750000	0.02083333	0.02083333	0.02083333	0.02083333	0.07083333

SITE

CUS_ID	SITE1	SITE2	SITE3	SITE4	SITE5	SITE6	SITE7	SITE8	SITE9	SITE10
1	1	0.01702786	0.10371517	0.007739938	0.06656347	0.013931889	0.01083591	0.01083591	0.08823529	0.12538700	0.03250774
2	2	0.03431373	0.03151261	0.327030812	0.06792717	0.004901961	0.04131653	0.03291317	0.02731092	0.02450980	0.02450980
3	3	0.14690027	0.09568733	0.216981132	0.07951482	0.025606469	0.01482480	0.06873315	0.05525606	0.05256065	0.01212938
4	4	0.08398438	0.15039062	0.033203125	0.03320312	0.017578125	0.09179688	0.19335938	0.05273438	0.01367188	0.03710938
5	5	0.16409692	0.05616740	0.007709251	0.08039648	0.027533040	0.03193833	0.01651982	0.02533040	0.05396476	0.15088106
6	6	0.03395062	0.15123457	0.015432099	0.03395062	0.052469136	0.05864198	0.04012346	0.01543210	0.02160494	0.07716049

	SITE11	SITE12	SITE13	SITE14	SITE15	SITE16	SITE17	SITE18	SITE19	SITE20
1	0.02321981	0.010835913	0.029411765	0.162538700	0.04798762	0.03560372	0.010835913	0.11609907	0.060371517	0.026315789
2	0.03571429	0.030112045	0.020308123	0.007703081	0.11974790	0.01610644	0.048319328	0.06232493	0.034313725	0.009103641
3	0.05256065	0.012129380	0.020215633	0.009433962	0.02021563	0.01212938	0.041778976	0.04716981	0.006738544	0.009433962
4	0.02148438	0.009765625	0.072265625	0.025390625	0.01757812	0.02148438	0.021484375	0.02929688	0.009765625	0.064453125
5	0.03193833	0.020925110	0.009911894	0.159691630	0.01872247	0.02312775	0.007709251	0.05176211	0.009911894	0.051762115
6	0.02777778	0.015432099	0.114197531	0.120370370	0.04012346	0.02160494	0.027777778	0.02777778	0.015432099	0.089506173

Feature Engineering_ClickStreams

clustering

	CUS_ID	category_cluster	category_cv	time_cluster	day_cluster	day_cv	month_cluster	month_cv	news_cluster	news_cv	game_cluster	game_cv	life_cluster
1	1	2	2.109661	2	3	0.23338501	4	1.5135520	4	1.5104795	3	3.624627	1
2	2	7	2.386097	1	1	0.70970025	1	0.5211001	1	1.5150108	3	2.656167	3
3	3	1	3.457935	1	1	0.60424375	4	1.0093134	2	0.8837821	2	0.000000	4
4	4	2	2.269863	2	1	0.55615702	4	1.4858977	4	1.0307252	2	3.872983	1
5	5	10	2.774812	1	3	0.44307890	4	0.7984766	1	1.6148339	3	2.774943	4
6	6	6	1.886650	1	3	0.48149792	1	0.7478754	1	1.7239483	2	0.000000	1
7	7	5	2.294393	2	3	0.18430023	1	0.6309688	4	1.4157538	1	3.872983	4
8	8	10	2.384441	1	1	0.69706503	2	0.5413901	3	0.8686207	2	0.000000	4
9	9	7	2.376340	2	3	0.22047336	1	0.2395940	4	0.9357081	1	2.119695	4
10	10	1	3.798568	1	3	0.25618534	4	1.0797259	4	1.2449835	1	2.839784	4
11	11	7	2.294687	2	2	0.28463478	3	1.1991187	3	0.5400210	3	3.391929	1
12	12	6	1.774923	1	2	0.69553448	1	0.4106160	5	0.9919575	3	3.345164	2
13	13	2	2.547496	1	3	0.40009535	3	0.9214798	4	1.5464084	2	0.000000	1
14	14	1	3.384532	1	1	0.66478163	3	0.9521600	5	1.6704993	2	0.000000	4
15	15	7	2.830148	1	3	0.24535741	2	0.6998147	3	1.6284087	2	0.000000	4

Tanimoto Similarity

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1.00000000	0.2297297	0.24666667	0.22222222	0.29943503	0.2049180	0.28658537	0.2258065	0.26699029	0.30000000	0.24456522	0.1965812	0.24000000
2	0.22972973	1.0000000	0.29906542	0.27142857	0.27888446	0.1666667	0.25311203	0.2657658	0.29889299	0.27777778	0.30081301	0.1379310	0.18840580
3	0.24666667	0.2990654	1.00000000	0.28368794	0.32768362	0.2063492	0.23428571	0.2913907	0.29126214	0.30674847	0.26486486	0.1983471	0.22137405
4	0.22222222	0.2714286	0.28368794	1.00000000	0.27272727	0.2368421	0.24242424	0.3237410	0.23786408	0.32026144	0.27428571	0.2181818	0.24166667
5	0.29943503	0.2788845	0.32768362	0.27272727	1.00000000	0.2187500	0.36315789	0.2727273	0.32051282	0.36170213	0.26484018	0.1604938	0.28481013
6	0.20491803	0.1666667	0.20634921	0.23684211	0.21875000	1.0000000	0.22222222	0.2109375	0.15306122	0.20138889	0.19753086	0.1931818	0.27659574
7	0.28658537	0.2531120	0.23428571	0.24242424	0.36315789	0.2222222	1.00000000	0.2732558	0.29464286	0.34659091	0.25242718	0.1901408	0.28671329
8	0.22580645	0.2657658	0.29139073	0.32374101	0.27272727	0.2109375	0.27325581	1.0000000	0.29951691	0.28571429	0.27419355	0.1746032	0.18978102
9	0.26699029	0.2988930	0.29126214	0.23786408	0.32051282	0.1530612	0.29464286	0.2995169	1.00000000	0.32870370	0.32758621	0.1711230	0.17587940
10	0.30000000	0.2777778	0.30674847	0.32026144	0.36170213	0.2013889	0.34659091	0.2857143	0.32870370	1.00000000	0.29441624	0.2296296	0.23972603
11	0.24456522	0.3008130	0.26486486	0.27428571	0.26484018	0.1975309	0.25242718	0.2741935	0.32758621	0.29441624	1.00000000	0.1614907	0.20238095
12	0.19658120	0.1379310	0.19834711	0.21818182	0.16049383	0.1931818	0.19014085	0.1746032	0.17112299	0.22962963	0.16149068	1.0000000	0.21505376
13	0.24000000	0.1884058	0.22137405	0.24166667	0.28481013	0.2765957	0.28671329	0.1897810	0.17587940	0.23972603	0.20238095	0.2150538	1.00000000
14	0.20348837	0.2735043	0.26946108	0.25625000	0.22009569	0.2112676	0.26203209	0.2647059	0.24890830	0.25268817	0.25742574	0.1619718	0.19205298
15	0.22543353	0.3407080	0.34782609	0.30379747	0.27450980	0.2123288	0.24226804	0.3095238	0.30493274	0.24607330	0.28217822	0.1333333	0.20129870

Feature Engineering_Searchkeywords

```
ts.dt$QRY_STR <- gsub('pre ₩₩d+', '', ts.dt$QRY_STR)
ts.dt$QRY_STR <- gsub('pre', '', ts.dt$QRY_STR)

ts.dt$QRY_STR <- gsub('qdt ₩₩d+', '', ts.dt$QRY_STR)
ts.dt$QRY_STR <- gsub('qdt', '', ts.dt$QRY_STR)

ts.dt$QRY_STR <- gsub('query ₩₩d+', '', ts.dt$QRY_STR)
ts.dt$QRY_STR <- gsub('query', '', ts.dt$QRY_STR)

ts.dt$QRY_STR <- gsub('sm ₩₩d+', '', ts.dt$QRY_STR)
ts.dt$QRY_STR <- gsub('sm', '', ts.dt$QRY_STR)

ts.dt$QRY_STR <- gsub('sug ₩₩d+', '', ts.dt$QRY_STR)
ts.dt$QRY_STR <- gsub('sug', '', ts.dt$QRY_STR)

ts.dt$QRY_STR <- gsub('top ₩₩d+', '', ts.dt$QRY_STR)
ts.dt$QRY_STR <- gsub('top', '', ts.dt$QRY_STR)

ts.dt$QRY_STR <- gsub('utf₩₩d+', '', ts.dt$QRY_STR)
ts.dt$QRY_STR <- gsub('utf', '', ts.dt$QRY_STR)
```

불용성 처리

Feature Engineering_Searchkeywords

라인교육 M40+ 서비스 사회/문화/종교 여행 커뮤니티 온라인교
포츠/레저 엔터테인먼트 정치/행정 문학/예술 인터넷/컴퓨터 M4
문학/예술 커뮤니티 서비스 뉴스/미디어 인터넷/컴퓨터 온라인
핑 정치/행정 교육/학원 게임 서비스 스포츠/레저 비즈니스/경제
라인교육 여행 사회/문화/종교 M40+ 여행 사회/문화/종교 금융/
인터넷/컴퓨터 뉴스/미디어 엔터테인먼트 게임 서비스 문학/예
엔터테인먼트 온라인교육 커뮤니티 게임 뉴스/미디어 비즈니스/경
핑 M40+ 엔터테인먼트 여행 쇼핑 정치/행정 뉴스/미디어 금융/부
교육 비즈니스/경제 사회/문화/종교 커뮤니티 게임 M40+ 교육/행
/행정 온라인교육 금융/부동산 여행 엔터테인먼트 사회/문화/종
인터넷/컴퓨터 교육/학원 스포츠/레저 정치/행정 쇼핑 금융/부동
임 여행 뉴스/미디어 M40+ 정치/행정 서비스 뉴스/미디어 게임 비
문학/예술 커뮤니티 엔터테인먼트 여행 쇼핑 온라인교육 사회/문
/부동산 여행 인터넷/컴퓨터 뉴스/미디어 사회/문화/종교 스포츠
스/미디어 서비스 비즈니스/경제 온라인교육 사회/문화/종교 커
테인먼트 스포츠/레저 쇼핑 문학/예술 여행 M40+ 사회/문화/종교
컴퓨터 커뮤니티 게임 여행 비즈니스/경제 엔터테인먼트 서비스
행정 교육/학원 여행 인터넷/컴퓨터 비즈니스/경제 사회/문화/종
온라인교육 M40+ 사회/문화/종교 정치/행정 금융/부동산 엔터테
경제 쇼핑 인터넷/컴퓨터 여행 서비스 온라인교육 뉴스/미디어
라인교육 유통/판매/운송 생활/가정/취미 제조 건강/의학 교육/
정보통신/IT 정치/행정 여행 F20- 제조 뉴스/미디어 엔터테인
금융/부동산 여행 유통/판매/운송 정치/행정 게임 서비스 사회

trans2vec

F20.	F30	F40.	M20.	M30	M40.
0.043344	0.471441	0.13739	0.006333	0.220689	0.120802
0.04185	0.432659	0.116728	0.00787	0.271956	0.128937
0.161492	0.094548	0.023701	0.269228	0.39154	0.059491
0.04045	0.230293	0.18212	0.024053	0.244958	0.278125
0.126633	0.418368	0.021534	0.017347	0.392911	0.023208
0.438311	0.138773	0.09643	0.202929	0.081113	0.042444
0.05164	0.163384	0.019496	0.04462	0.649885	0.070975
0.285774	0.529205	0.037667	0.012997	0.123205	0.011153
0.073982	0.334478	0.215267	0.021109	0.180211	0.174953

Feature Engineering_CosineSimilarity

```
cs.dt <- fread("train_profiles.csv")
cs.dt$GENDER<-substr(cs.dt$GROUP, 1, 1)
tr.dt <- fread("train_clickstreams.tab"); tr.dt[,CUS_ID:= as.numeric(CUS_ID)]
ts.dt <- fread("test_clickstreams.tab"); ts.dt[,CUS_ID:= as.numeric(CUS_ID)]
setkey(cs.dt, CUS_ID); setkey(tr.dt, CUS_ID) ; setkey(ts.dt, CUS_ID)
md.dt <- merge(cs.dt, tr.dt)
head(md.dt)
md.dt$GENDER<-substr(md.dt$GROUP, 1, 1)
###### Make sites sentences
f <- function(x, t) {
  grp <- md.dt[CUS_ID==x, GENDER][1]
  itemfreq <- table(md.dt[CUS_ID==x,  ACT_NM])
  fitems <- itemfreq[itemfreq >= t]
  act <- names(fitems)
  #
  sapply(act, function(x) gsub(" ", "_", x))
  set.seed(1)
  #
  as.vector((sapply(1:20, function(x) c(grp, sample(act)))))
}
items <- unlist(sapply(cs.dt$CUS_ID, f, 2))
write.table(items, "items.txt", eol = " ", quote = F, row.names = F, col.names = F)

##### Train site2vec model
set.seed(12345)
model = train_word2vec("items.txt","vec.bin",vectors=300,threads=1,window=5,cbow=1,iter=5,negative_samples=10, force = T)
model <- read.binary.vectors("vec.bin") # reload the model.

##### Explore the model
for (v in unique(md.dt[,GENDER])) print(closest_to(model, v, n=10))
model[[unique(md.dt[,GENDER]), average=F]] %>% plot(method="pca")
items.1 <- c(unique(md.dt[,GENDER]), unique(md.dt[CUS_ID==1, ACT_NM]))
model[[items.1[1:10], average=F]] %>% plot(method="pca")

##train cosine 유사도
cosineSimilarity(model[[unique(md.dt[CUS_ID==1, ACT_NM]), average=T]], model[[c("M","F"), average=F]])
cosineSimilarity(model[[unique(md.dt[CUS_ID==2, ACT_NM]), average=T]], model[[c("M","F"), average=F]])

##test cosine 유사도
cosineSimilarity(model[[unique(ts.dt[CUS_ID==2501, ACT_NM]), average=T]], model[[c("M","F"), average=F]])
cosineSimilarity(model[[unique(ts.dt[CUS_ID==2502, ACT_NM]), average=T]], model[[c("M","F"), average=F]])

#### train _ word2vec
result=NULL
for (i in 1:2500){
  a <- i
  b <- cosineSimilarity(model[[unique(md.dt[CUS_ID==i, ACT_NM]), average=T]], model[[c("M","F"), average=F]])
  c <- data.frame(a,b)
  result=rbind(result,c)
```

CosineSimilarity

ACT_NM,SITE_NM,SITE,
MACT_NM, QRY_STR merge

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
1	0.09600750	0.04156002	0.101570214	0.01646666	0.008948699	0.15191374	0.04940456	0.01299943	0.04201560	0.06311563	0.01197886	0.37967154	0.2531469
2	0.08305699	0.08968881	0.004249424	0.13466007	0.136515082	0.03484446	0.13212549	0.13212886	0.02217125	0.16756991	0.13216511	0.08607127	0.2437319
3	0.04820711	0.17272038	0.055157726	0.10733353	0.167186481	0.03671107	0.04787565	0.18906094	0.08247310	0.19711873	0.11754664	0.05635807	0.1838156
4	0.10575573	0.05771229	0.049305862	0.11889285	0.085968877	0.04316248	0.10359547	0.05916272	0.03927565	0.08258394	0.09705969	0.16675820	0.3135750
5	0.14176174	0.05753706	0.176602126	0.04737598	0.065542435	0.18915014	0.06110262	0.02552243	0.11378523	0.07409996	0.07619798	0.35920928	0.2150626
6	0.10257981	0.08664485	0.167713691	0.05264577	0.005652529	0.16913095	0.05270011	0.04624910	0.13642028	0.05259950	0.02768118	0.30677188	0.3017352

	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25	X26	X27
1	0.06328611	0.08881732	0.1107837	0.09135338	0.3930788	0.1507235	0.3907377	0.2171736	0.1474807	0.3498551	0.1417847	0.33058460	0.2046519	0.04638094
2	0.24681876	0.12795773	0.3470557	0.19271523	0.2262546	0.2984944	0.1264497	0.2903189	0.2911798	0.1345570	0.2063151	0.07321418	0.2302472	0.20499838
3	0.18038480	0.14321759	0.3271427	0.29572192	0.2110737	0.3242169	0.1115548	0.2717478	0.3831652	0.1688579	0.2764233	0.13421154	0.2257514	0.27912347
4	0.13610833	0.13810272	0.2383770	0.27935763	0.3600451	0.2673347	0.1921097	0.3199161	0.3331145	0.3142928	0.2460723	0.14059445	0.3243542	0.22982417
5	0.10828938	0.15079245	0.1799054	0.17656029	0.3751601	0.2208209	0.3323110	0.2450448	0.2273033	0.3214339	0.1771279	0.28889867	0.2228348	0.10534526
6	0.15368085	0.12361478	0.1265655	0.13640542	0.3753423	0.2357644	0.2922423	0.3436529	0.1828015	0.3935465	0.2265980	0.32823395	0.3152397	0.09360274

	X28	X29	X30	X31	X32	X33	X34	X35	X36	X37	X38	X39	X40	X41
1	0.34633432	0.1573240	-0.003161647	0.11206811	0.07560236	0.02489578	0.1111393	0.1541025	0.2426759	0.2846867	0.2387240	0.1446443	0.2294572	0.2736157
2	0.05071457	0.1520089	0.155662922	0.08905944	0.23578190	0.09013498	0.2205315	0.2841002	0.2392778	0.2858123	0.3594905	0.2207950	0.3059842	0.3676447
3	0.05815388	0.1296463	0.142770209	0.11743148	0.30154476	0.18521747	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
4	0.11058066	0.2393170	0.115206056	0.10093400	0.19660638	0.24059101	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
5	0.32282481	0.1665351	0.044848476	0.14944507	0.12544326	0.06430882	0.2445419	0.2148339	0.1822077	0.2224575	0.2471978	0.1961816	0.3025358	0.2837076
6	0.30145264	0.2455017	0.117148423	0.19444359	0.03716538	0.11558590	0.2257780	0.1527922	0.2073141	0.2477832	0.2157057	0.1983494	0.2347589	0.2921783

	X42	X43	X44	X45	X46	X47	X48	X49	X50	X51	X52	X53	X54
1	0.3400553	0.1902952	0.2956560	0.04009222	0.017713176	0.05091890	0.03038236	0.04496998	0.06821000	0.03021596	0.02018127	0.01773963	0.045655113
2	0.3450689	0.3662442	0.4011376	0.06729354	0.075798892	0.03176744	0.09667985	0.10817985	0.07949845	0.10749025	0.13583419	0.03421338	0.129344248
3	0.0000000	0.0000000	0.0000000	-0.01235308	0.020250413	-0.00435375	0.02718295	0.01832060	0.01567136	0.01228487	0.06027945	0.02323227	0.015541231
4	0.0000000	0.0000000	0.0000000	0.05050842	0.009093642	0.02295943	0.05687421	0.04458409	0.06455856	0.06193355	0.03071157	-0.01078604	0.015967413
5	0.3283350	0.2255045	0.3076531	0.04824912	0.020670277	0.04639823	0.05266025	0.05952985	0.08788576	0.06518724	0.05099418	0.01884388	0.064596721
6	0.2944299	0.1154954	0.3253614	0.01863364	0.015507898	0.06641464	-0.03650555	-0.01697636	0.07741083	-0.02562744	-0.01529943	0.07987721	-0.003438269

	X55
1	0.074368498
2	0.120539086
3	0.017216221
4	0.081796100
5	0.089805745
6	0.004579682

Model evaluation

2500개의 Train data ✚ 2500개의 Test data 〓 Test_public을 이용하여 logloss 확인

3503개의 Train data ✚ 1497개의 Test data 〓 최종 submission 생성