Popularity_youku_151205 160103 · last edited by 郭鹏 13 days ago

Page History Edit Delete

Video Property

序号	特征	说明
1	Duration	视频时长，数值型，直接从metadata提取。
2	Publication Time	发布时间(保留小时)，数值型，直接从metadata提取。
3	Category	视频类别，类别型，直接从metadata提取。能转义为25维矢量。
4	Public Type	公开类型，类别型，直接从metadata提取。能转义为4维矢量。
5	Copyright Type	版权所有，类别型，直接从metadata提取。能转义为2维矢量。
6	State	视频状态，类别型，直接从metadata提取。能转义为2维矢量。
7	Operation Limit	操作限制，类别型，直接从metadata提取。能转义为2维矢量。
8	Stream Types	视频格式，类别型，直接从metadata提取。能转义为10维矢量。
9	Source	上传来源(取前10及其他)，类别型，直接从metadata提取。能转义为11维矢量。
10	Tag Count	tag数目，数值型，直接从metadata提取。

User Statistic

序号	特征	说明
1	Following Count	关注的用户数，数值型，直接从metadata提取。
2	Regist Time	注册时间，数值型，直接从metadata提取。
3	Favorites Count	收藏数，数值型，直接从metadata提取。
4	Followers Count	被关注的用户数，数值型，直接从metadata提取。
5	Statuses Count	状态数，数值型，直接从metadata提取。
6	Playlists Count	播放列表数目，数值型，直接从metadata提取。
7	Subscribe Count	订阅数，数值型，直接从metadata提取。
8	Videos Count	发布的视频数，数值型，直接从metadata提取。
9	VV Count	发布的视频的总播放数，数值型，直接从metadata提取。
10	Is Vip	是否为vip，bool型，直接从metadata提取。
11	Is Verified	是否认证，bool型，直接从metadata提取。
12	Is Share	是否分享，bool型，直接从metadata提取。
13	Gender	用户性别，类别型，直接从metadata提取。能转义为3维矢量。

Historical Popularity

序号	特征	说明
1-7	Daily Increase of View Count Day1-Day7	第1-7天播放量增量，数值型型。从metadata提取vc，后项减前项。
8-14	Growth Rate of View Count Day1-Day7	第1-7天播放量增长率，数值型型。各天播放量增量/七天总播放量。
15	Favorite_count	前7天播放数，数值型，直接从metadata提取。
16	Favorite_count	前7天收藏数，数值型，直接从metadata提取。
17	Up_count	前7天顶数，数值型，直接从metadata提取。
18	Comment_count	前7天评论数，数值型，直接从metadata提取。
19	Down_count	前7天踩数，数值型，直接从metadata提取。

Topic

序号	特征	说明
1	Tag Topic	给top50的tag的归一化的分数值，数值型，分词并统计词频后归一化得到。
2	Title Topic	给top50的title的归一化的分数值，数值型，分词并统计词频后归一化得到。

Natural Language Processing

序号	特征	说明
1	sentiment_tag	tag为正面情绪的概率，数值型，由第三方库SnowNLP得到。

2	sentiment_title	title为正面情绪的概率，数值型，由第三方库SnowNLP得到。
3	sentiment_description	description为正面情绪的概率，数值型，由第三方库SnowNLP得到。
4	title_word_property	title所含词的词性，类别型，由第三方库jieba得到。能转义为56维矢量。
5	description_word_property	description所含词的词性，类别型，由第三方库jieba得到。能转义为58维矢量。
6	title_length	title长度，数值型，直接从metadata提取。
7	title_chinese	title中的中文数，数值型，直接从metadata提取。
8	title_chinese_rate	title中的中文比例，数值型，计数 / 总长。
9	title_non_chinese	title中的非中文数，数值型，直接从metadata提取。
10	title_non_chinese_rate	title中的非中文比例，数值型，计数 / 总长。
11	title_character	title中的字母数，数值型，直接从metadata提取。
12	title_character_rate	title中的字母比例，数值型，计数 / 总长。
13	title_number	title中的数字数，数值型，直接从metadata提取。
14	title_number_rate	title中的数字比例，数值型，计数 / 总长。
15	title_symbol	title中的标点数，数值型，直接从metadata提取。
16	title_symbol_rate	title中的标点比例，数值型，计数 / 总长。
17	description_length	description长度，数值型，直接从metadata提取。
18	name_length	用户名长度，数值型，直接从metadata提取。
19	name_chinese	name中的中文数，数值型，直接从metadata提取。
20	name_chinese_rate	name中的中文比例，数值型，计数 / 总长。
21	name_non_chinese	name中的非中文数，数值型，直接从metadata提取。
22	name_non_chinese_rate	name中的非中文比例，数值型，计数 / 总长。
23	name_character	name中的字母数，数值型，直接从metadata提取。
24	name_character_rate	name中的字母比例，数值型，计数 / 总长。
25	name_number	name中的数字数，数值型，直接从metadata提取。
26	name_number_rate	name中的数字比例，数值型，计数 / 总长。
27	name_symbol	name中的标点数，数值型，直接从metadata提取。
28	name_symbol_rate	name中的标点比例，数值型，计数 / 总长。

...

Results

📎 feature_importance_all 📎 feature_importance-nlp📎 feature_importance-topic📎 feature_importance-user_statistic 📎 feature_importance-video_property
📎 feature_importance_round_4

📎 结果 📎 代码

模型参数选取及第一轮结果（数据集均分为1：1的训练集和测试集）

方法	可调参数	搜索范围	最终参数取值	调前准确率L1-L5，Macro	调前召回率L1-L5，Macro	调前F1值L1-L5，Macro	调后准确率L1-L5，Macro	调后召回率L1-L5，Macro	调后F1值L1-L5，Macro
KNN	n_neighbors	[1,100]	7	72.3983%，53.5154%，43.5765%，50.8279%，64.8746%，Marco 57.03854%	78.8686%，59.669%，20.863%，18.0778%，18.413%，Marco 39.17828%	75.4950694882%，56.4249207947%，28.2167465452%，26.6699739099%，28.6846063472%，Marco 43.098263417%	72.5235%，62.3433%，55.0808%，62.4796%，78.3898%，Marco 66.1634%	87.2527%，58.7608%，27.3479%，22.8832%，31.3666%，Marco 45.52224%	79.2091837013%，60.4990612645%，36.5490226176%，33.4978042595%，44.8050683273%，Marco 50.912028034%
decision_tree	max_depth，criterion	[1, 40]，["gini"，"entropy"]	17, gini	79.2506%，61.113%，41.6717%，45.964%，52.5274%，Marco 56.10534%	79.2436%，60.7909%，42.7007%，45.7835%，50.39%，Marco 55.78174%	79.2470998454%，60.9515244664%，42.1799251933%，45.8735724461%，51.4365051196%，Marco 55.9377254142%	78.3137%，65.2744%，52.5976%，57.361%，66.4194%，Marco 63.99322%	86.1825%，63.8067%，38.11%，40.4063%，45.5409%，Marco 54.80928%	82.0598950037%，64.5322058532%，44.196837663%，47.4135170819%，54.0334253027%，Marco 58.4471761809%
gradient_boosting	n_estimators，max_depth，max_features	[100,1000]，[1, 30]，[1, 200]	415, 5, 36	78.8065%，64.4944%，55.2293%，60.0331%，71.9734%，Marco 66.10734%	87.0543%，66.6941%，32.772%，30.9747%，40.3188%，Marco 51.56278%	82.7253298302%，65.5758082917%，41.1351791303%，40.8647887889%，51.6844646364%，Marco 56.3971141355%	81.9114%，67.3536%，59.5103%，65.5776%，73.1163%，Marco 69.49384%	87.2274%，70.966%，39.5117%，40.944%，53.3062%，Marco 58.39106%	84.4858595705%，69.1126286889%，47.4915295694%，50.4124844989%，61.6591526201%，Marco 62.6323309896%
naive_bayes	none	none	none	50.9905%，11.8402%，24.2791%，37.3901%	99.8146%，0.244%，0.8989%，4.72%	67.4989952104%，0.478146472253%，1.73361529828%，8.3818975495%	50.9905%，11.8402%，24.2791%，37.3901%	99.8146%，0.244%，0.8989%，4.72%	67.4989952104%，0.478146472253%，1.73361529828%，8.3818975495%

				44.9324%, Marco 33.88646%	18.04%, Marco 24.7435%	25.7439924792%, Marco 20.7673294019%	44.9324%, Marco 33.88646%	18.04%, Marco 24.7435%	25.7439924792%, Marco 20.7673294019%
random_forests	n_estimators, max_depth, max_features	[300, 1000], [10, 30], [50,200]	608, 21, 91	80.9678%, 69.1872%, 63.4302%, 71.4246%, 84.6604%, Marco 73.93404%	89.3864%, 69.7275%, 40.9478%, 44.2472%, 49.0336%, Marco 58.6685%	84.9690839195%, 69.4562992685%, 49.7677124214%, 54.643198448%, 62.1000821195%, Marco 64.1872752354%	81.3957%, 69.5538%, 63.9374%, 71.8457%, 83.3152%, Marco 74.00956%	89.3207%, 70.3879%, 41.9534%, 45.1519%, 51.8142%, Marco 59.72562%	85.1742527489%, 69.9683642405%, 50.6633497369%, 55.4536137806%, 63.8929860688%, Marco 65.0305133151%
svm_linear	C	$10^{[0,10]}$	6.546176289194936	77.4603%, 57.3778%, 3.0776%, 11.1787%, 0.0%, Marco 29.81888%	52.1162%, 30.0795%, 5.1624%, 87.0946%, 0.0%, Marco 34.89054%	62.3097010162%, 39.4683013333%, 3.85626267961%, 19.8142202413%, 0%, Marco 25.0896970541%	71.8286%, 61.2664%, 42.3486%, 70.4861%, 31.6779%, Marco 55.52152%	86.5725%, 54.2263%, 30.1546%, 1.7327%, 48.0163%, Marco 44.14048%	78.514372356%, 57.5317779621%, 35.2261719086%, 3.38225684919%, 38.1723023701%, Marco 42.5653762892%
svm_non_linear	C, gamma	[1,10], [0, 1e-3]	5.199814870749871, 5.238333199546759e-06	50.075%, 69.0909%, 42.8571%, 100.0%, 0.0%, Marco 52.4046%	99.7808%, 1.0558%, 0.3132%, 0.7812%, 0.0%, Marco 20.3862%	66.6844200892%, 2.07981764559%, 0.62185547564%, 1.55028914123%, 0%, Marco 14.1872764703%	76.6554%, 50.0043%, 37.7237%, 44.4444%, 49.2537%, Marco 51.6163%	74.0257%, 67.7211%, 11.208%, 11.0775%, 12.5475%, Marco 35.31596%	75.3176031205%, 57.5295764674%, 17.2815262744%, 17.734725973%, 19.9999611901%, Marco 37.5726786051%
xgboost	n_estimators, subsample, learning_rate, max_depth, colsample_bytree	[100, 1000], [0.7, 1], [0, 0.1], [1, 20], [0.1, 1]	745, 0.9153215936899797, 0.015241899846790053, 10, 0.24207133023642236	76.8211%, 63.1147%, 56.028%, 58.9358%, 72.5282%, Marco 65.48556%	88.0205%, 64.3181%, 25.6397%, 28.4568%, 37.0634%, Marco 48.6997%	82.0403542862%, 63.710717901%, 35.1801535148%, 38.381379509%, 49.0574403126%, Marco 53.6740091047%	80.9112%, 67.9989%, 63.121%, 72.6889%, 85.9941%, Marco 74.14282%	88.8211%, 70.2615%, 38.4613%, 41.2086%, 49.1353%, Marco 57.57756%	84.6818405963%, 69.1116865328%, 47.7980064893%, 52.5983064517%, 62.5377734487%, Marco 63.3455227038%
ols	none	none		52.858%, 33.313%, 7.5904%, 4.7523%, 15.5124%, Marco 22.80522%	53.8202%, 1.6029%, 9.5705%, 37.2738%, 50.9664%, Marco 30.64676%	53.3347606465%, 3.05862989068%, 8.46621368343%, 8.4298224075%, 23.7853626528%, Marco 19.4149578562%			
ensemble	none	none		77.7283%, 70.6863%, 61.716%, 76.0716%, 87.125%, Marco 74.66544%	91.995%, 64.1525%, 40.5276%, 36.6593%, 47.2703%, Marco 56.12094%	84.2620307112%, 67.2610978554%, 48.9263163973%, 49.4759042264%, 61.2882278993%, Marco 62.2427154179%			

- 第一轮算法运行时间(minute)

方法	fit	predict
decision_tree	0.380300	0.010817
gradient_boosting	44.385633	0.659067
knn_norm	11.316850	666.770367
naive_bayes	0.048450	0.105317
random_forests	40.365167	8.306433
svm_linear	13.716867	0.006983
svm_non_linear	79.511367	10.898550
xgboost	46.400017	2.976600

第三轮结果（使用由一二轮确定的训练集及模型参数，训练集、测试集每次减少一组特征）

特征	方法	准确率L1-L5，Macro	召回率L1-L5，Macro	F1值L1-L5，Macro
-L	decision_tree	78.3308%, 66.0581%, 53.9704%, 58.719%, 66.9079%, Marco 64.79724%	86.838%, 63.7753%, 39.3705%, 40.6111%, 47.0329%, Marco 55.52556%	82.3653136718%, 64.8966312972%, 45.5286296404%, 48.0145128395%, 55.2369751645%, Marco 59.2084125227%
-L	gradient_boosting	81.6412%, 67.6444%, 59.4653%, 64.7299%, 71.577%, Marco 69.01156%	87.3889%, 70.3174%, 40.0592%, 42.028%, 53.7131%, Marco 58.70132%	84.4173275964%, 68.9550054082%, 47.8702700493%, 50.9651882849%, 61.3715298926%, Marco 62.7158642463%
-L	naive_bayes	50.9905%, 11.8402%, 24.2791%, 37.3901%, 44.9324%, Marco 33.88646%	99.8146%, 0.244%, 0.8989%, 4.72%, 18.04%, Marco 24.7435%	67.4989952104%, 0.478146472253%, 1.73361529828%, 8.38189754952%, 25.7439924792%, Marco 20.7673294019%
-L	random_forests	81.3042%, 69.3439%, 61.852%, 67.6015%, 77.3492%, Marco 71.49016%	88.7212%, 69.3493%, 42.9004%, 47.3199%, 53.0349%, Marco 60.26514%	84.8509244976%, 69.3465998949%, 50.6618567365%, 55.6710276737%, 62.9249592102%, Marco 64.6910736026%
-L	xgboost	80.2361%, 67.2997%, 62.2501%, 71.7874%, 83.7976%, Marco 73.07418%	88.6186%, 69.2376%, 36.8909%, 40.0051%, 49.983%, Marco 56.94704%	84.2192826313%, 68.2548975074%, 46.3271948859%, 51.3784398012%, 62.6167836114%, Marco 62.5593196874%
-L	knn	74.4537%, 62.8559%, 53.3249%,	85.1663%, 59.9453%, 34.6868%,	79.4505218683%, 61.3661068828%, 42.0323693627%,

		58.5563%, 72.7046%, Marco 64.37908%	34.551%, 39.471%, Marco 50.76408%	43.4590783172%, 51.1648391736%, Marco 55.4945831209%
-L	svm_linear	71.1331%, 60.3739%, 43.7043%, 52.4686%, 73.7546%, Marco 60.2869%	86.7395%, 55.0915%, 26.3698%, 9.9778%, 20.5832%, Marco 39.75236%	78.1649194027%, 57.6118683493%, 32.8929989865%, 16.7670577353%, 32.1844622775%, Marco 43.5242613503%
-L	svm_non_linear	50.2117%, 0.0%, 0.0%, 0.0%, 0.0%, Marco 10.04234%	100.0%, 0.0%, 0.0%, 0.0%, 0.0%, Marco 20.0%	66.8545792372%, 0%, 0%, 0%, 0%, Marco 13.3709158474%
-L	ols	54.7146%, 31.1395%, 7.0388%, 6.0839%, 18.1388%, Marco 23.42312%	55.0834%, 2.3133%, 12.5254%, 40.2441%, 49.3727%, Marco 31.90778%	54.8983806197%, 4.30666523281%, 9.01276673925%, 10.5698963905%, 26.5306364326%, Marco 21.063669083%
-T	decision_tree	78.4149%, 65.5117%, 52.7704%, 58.2266%, 66.8473%, Marco 64.35418%	86.2235%, 64.0026%, 38.8022%, 40.3551%, 46.0156%, Marco 55.0798%	82.1340237776%, 64.748357987%, 44.7209670771%, 47.6709219999%, 54.5089416962%, Marco 58.7566425075%
-T	gradient_boosting	81.7971%, 67.2341%, 59.4996%, 64.8871%, 72.4788%, Marco 69.17934%	87.2137%, 70.8974%, 38.816%, 40.9781%, 52.1533%, Marco 58.0117%	84.418602128%, 69.0171739442%, 46.9820958953%, 50.2327501863%, 60.6586681929%, Marco 62.2618580693%
-T	naive_bayes	50.9905%, 11.8402%, 24.2791%, 37.3901%, 44.9324%, Marco 33.88646%	99.8146%, 0.244%, 0.8989%, 4.72%, 18.04%, Marco 24.7435%	67.4989952104%, 0.478146472253%, 1.73361529828%, 8.38189754952%, 25.7439924792%, Marco 20.7673294019%
-T	random_forests	81.5751%, 69.4846%, 63.7572%, 71.8192%, 83.2421%, Marco 73.97564%	89.2044%, 70.6025%, 42.1118%, 45.2885%, 51.5429%, Marco 59.75002%	85.2193366351%, 70.0390895593%, 50.7217496143%, 55.5485905572%, 63.6649365447%, Marco 65.0387405821%
-T	xgboost	80.8733%, 67.819%, 62.7598%, 72.4346%, 86.723%, Marco 74.12194%	88.7075%, 70.2184%, 38.172%, 40.9099%, 48.7284%, Marco 57.34724%	84.6094399808%, 68.9978465199%, 47.4710068 7%, 52.2882405858%, 62.3968897066%, Marco 63.1526847326%
-T	knn	72.81%, 61.9584%, 53.2555%, 58.3735%, 76.7152%, Marco 64.62252%	85.7294%, 58.3081%, 30.3096%, 28.9775%, 37.5381%, Marco 48.17254%	78.7432980571%, 60.0778534844%, 38.632225721%, 38.7292211022%, 50.4097973384%, Marco 53.3184791406%
-T	svm_linear	71.1596%, 59.5693%, 48.487%, 41.2374%, 78.174%, Marco 59.72546%	86.4274%, 56.5915%, 21.3555%, 13.9382%, 18.5826%, Marco 39.37904%	78.0538903976%, 58.0422318192%, 29.6514050471%, 20.8343952283%, 30.027433217%, Marco 43.3218711418%
-T	svm_non_linear	50.2117%, 0.0%, 0.0%, 0.0%, 0.0%, Marco 10.04234%	100.0%, 0.0%, 0.0%, 0.0%, 0.0%, Marco 20.0%	66.8545792372%, 0%, 0%, 0%, 0%, Marco 13.3709158474%
-T	ols	53.3185%, 32.4093%, 8.1476%, 5.5307%, 19.487%, Marco 23.77862%	52.314%, 1.7411%, 10.862%, 46.3213%, 40.963%, Marco 30.44028%	52.8114739119%, 3.30466596175%, 9.31100404006%, 9.88155573208%, 26.4101234409%, Marco 20.3437646173%
-U	decision_tree	71.1468%, 62.9866%, 48.1072%, 51.9916%, 64.7811%, Marco 59.80266%	87.3725%, 54.7534%, 28.9286%, 21.1676%, 32.6212%, Marco 44.96866%	78.429235847%, 58.5821386859%, 36.1305768466%, 30.0860969546%, 43.3919367268%, Marco 49.3239970122%
-U	gradient_boosting	74.8913%, 62.0468%, 54.996%, 66.8971%, 71.2657%, Marco 66.01938%	86.127%, 62.6231%, 28.6221%, 23.1308%, 40.2848%, Marco 48.15756%	80.1171419037%, 62.3336179957%, 37.6497674929%, 34.3756422327%, 51.4730901495%, Marco 53.1898519549%
-U	naive_bayes	74.6448%, 52.9922%, 26.3685%, 17.7119%, 9.7427%, Marco 36.29202%	68.8565%, 53.028%, 22.5781%, 26.4681%, 41.6073%, Marco 42.5076%	71.633910929%, 53.0100939557%, 24.3265366685%, 21.2222879307%, 15.7884105827%, Marco 37.1962480133%
-U	random_forests	73.0128%, 65.521%, 58.9252%, 75.1872%, 86.8729%, Marco 71.90382%	89.2954%, 59.2458%, 31.8318%, 25.7084%, 35.2323%, Marco 48.26274%	80.3373727405%, 62.2255930552%, 41.3344465189%, 38.3156968684%, 50.1327064641%, Marco 54.4691631294%
-U	xgboost	74.8581%, 62.8796%, 58.3886%, 76.8964%, 86.5956%, Marco 71.92366%	87.3526%, 62.6535%, 29.4004%, 24.5732%, 38.5554%, Marco 48.50702%	80.6241470638%, 62.7663463835%, 39.1085032394%, 37.2444676333%, 53.3551948644%, Marco 54.6197318369%
-U	knn	72.093%, 60.9611%, 50.1624%, 54.5906%, 73.0797%, Marco 62.17736%	85.1704%, 57.538%, 28.1916%, 24.4111%, 33.8759%, Marco 45.8374%	78.087967667%, 59.2001082169%, 36.0966463956%, 33.7364030308%, 46.2928656233%, Marco 50.6827981867%
-U	svm_linear	92.0545%, 79.2418%, 11.1723%, 58.2126%, 78.5937%, Marco 63.85498%	20.4232%, 1.6794%, 98.4399%, 4.114%, 17.0566%, Marco 28.34262%	33.4296925417%, 3.28909306634%, 20.0671110473%, 7.68489333286%, 28.0300491147%, Marco 18.5001678206%
-U	svm_non_linear	50.2117%, 0.0%, 0.0%, 0.0%, 0.0%, Marco 10.04234%	100.0%, 0.0%, 0.0%, 0.0%, 0.0%, Marco 20.0%	66.8545792372%, 0%, 0%, 0%, 0%, Marco 13.3709158474%
-U	ols	51.2396%, 32.4665%, 7.447%, 3.7871%, 12.5286%, Marco 21.49376%	52.457%, 1.6392%, 8.8611%, 29.353%, 50.0509%, Marco 28.47224%	51.8411538508%, 3.12083240045%, 8.09274062582%, 6.70865485017%, 20.0406748453%, Marco 17.9608113145%
-V	decision_tree	78.7751%, 59.6604%, 51.8331%, 58.3607%, 63.3705%, Marco 62.39996%	78.9616%, 66.8871%, 37.3455%, 39.4418%, 46.2869%, Marco 53.78458%	78.8682397459%, 63.0674037945%, 43.4125011169%, 47.0714155008%, 53.4979672407%, Marco 57.1835054797%
-V	gradient_boosting	79.8488%, 62.5309%, 58.0812%, 62.8877%, 67.1972%, Marco 66.10916%	83.0983%, 68.4352%, 36.7566%, 40.1502%, 51.5429%, Marco 55.99664%	81.4411491464%, 65.3499592288%, 45.0214457931%, 49.0101939682%, 58.3381445675%, Marco 59.8321785408%
-V	naive_bayes	50.9905%, 11.8402%, 24.2791%, 37.3901%, 44.9324%, Marco 33.88646%	99.8146%, 0.244%, 0.8989%, 4.72%, 18.04%, Marco 24.7435%	67.4989952104%, 0.478146472253%, 1.73361529828%, 8.38189754952%, 25.7439924792%, Marco 20.7673294019%
-V	random_forests	82.0962%, 64.9914%, 62.4902%, 69.1254%, 78.016%, Marco 71.34384%	84.0563%, 72.1456%, 41.2405%, 46.0737%, 53.0688%, Marco 59.31698%	83.0646883563%, 68.3818888825%, 49.6888017356%, 55.2931913874%, 63.1685062006%, Marco 63.9194153125%
-V	xgboost	79.1375%, 63.6605%, 62.1463%, 70.4987%, 83.0111%, Marco 71.69082%	85.466%, 67.7777%, 36.5361%, 39.9368%, 48.0502%, Marco 55.55336%	82.1800942872%, 65.6546159465%, 46.0180018206%, 50.9889027018%, 60.8677001864%, Marco 61.1418629885%
-V	knn	64.0182%, 52.5874%, 50.7205%, 56.6608%, 71.3339%, Marco 59.06416%	82.0479%, 43.3419%, 24.7271%, 24.7952%, 29.1963%, Marco 40.82168%	71.9203000803%, 47.5191173512%, 33.2461436958%, 34.4950861363%, 41.4340356345%, Marco 45.7229365796%
-V	svm_linear	76.7051%, 39.7715%, 69.2308%, 56.6406%, 94.4223%, Marco 67.35406%	35.9782%, 86.5867%, 0.9298%, 2.4752%, 8.0366%, Marco 26.8013%	48.9817289487%, 54.5066792507%, 1.83495573983%, 4.74312495543%, 14.8124615076%, Marco 24.9757900804%
-V	svm_non_linear	50.2117%, 0.0%, 0.0%, 0.0%, 0.0%, Marco 10.04234%	100.0%, 0.0%, 0.0%, 0.0%, 0.0%, Marco 20.0%	66.8545792372%, 0%, 0%, 0%, 0%, Marco 13.3709158474%
-V	ols	52.3689%, 32.9918%, 6.4303%, 5.8074%, 13.137%, Marco 22.14708%	48.7186%, 2.0193%, 9.9356%, 46.7139%, 47.2024%, Marco 30.91796%	50.4778432851%, 3.80566972989%, 7.80756190371%, 10.3305250578%, 20.5536657242%, Marco 18.5950531401%

-V: declude video property features
-U: declude user statistic features
-T: declude content topic features
-L: declude natural language features

第四轮结果（使用由一二轮确定的训练集及模型参数，训练集、测试集增加历史播放量特征）

- 使用1天播放记录

方法	准确率L1-L5， Macro	召回率L1-L5， Macro	F1值L1-L5， Macro
decision_tree	89.5368%, 82.4181%, 70.9408%, 72.3311%, 77.2147%, Marco 78.4883%	93.8104%, 80.5875%, 63.4673%, 64.8856%, 65.6155%, Marco 73.67326%	91.6237937936%, 81.4925209165%, 66.9962753114%, 68.4063502789%, 70.9441161302%, Marco 75.8926112861%
gradient_boosting	90.3873%, 85.7912%, 79.0468%, 82.0349%, 85.2953%, Marco 84.5111%	95.9448%, 83.473%, 67.8582%, 70.1946%, 72.9739%, Marco 78.0889%	93.0831716171%, 84.6162252573%, 73.0264261088%, 75.6542850307%, 78.6549839472%, Marco 81.0070183922%
naive_bayes	51.0196%, 12.0581%, 23.2346%, 44.9919%, 55.6358%, Marco 37.388%	99.8146%, 0.244%, 1.0538%, 4.7542%, 26.1105%, Marco 26.39542%	67.5244867034%, 0.478321002105%, 2.01615762916%, 8.59968885923%, 35.5411451258%, Marco 22.8319598639%
random_forests	90.0197%, 85.5341%, 79.0124%, 82.5257%, 92.8405%, Marco 85.98648%	96.1453%, 83.2554%, 66.9009%, 68.607%, 67.7179%, Marco 76.5253%	92.9817211872%, 84.3793684932%, 72.4539938602%, 74.9254224916%, 78.3137312648%, Marco 80.6108474594%
xgboost	90.2591%, 85.7796%, 79.2805%, 83.8221%, 94.3934%, Marco 86.70694%	96.1775%, 83.6885%, 67.8514%, 68.5473%, 69.6507%, Marco 77.18308%	93.1243606701%, 84.7211487542%, 73.1220478727%, 75.4190622964%, 80.1560846794%, Marco 81.3085408546%
svm_linear	75.5849%, 63.0871%, 40.5157%, 32.1353%, 58.3127%, Marco 53.92714%	82.3797%, 54.774%, 37.8758%, 37.3933%, 63.9878%, Marco 55.28212%	78.8361618556%, 58.6373759519%, 39.1512995684%, 34.5654856704%, 61.0185794017%, Marco 54.4417804896%
knn	72.7942%, 62.2576%, 53.8715%, 61.5331%, 84.0241%, Marco 66.8961%	85.945%, 58.2591%, 30.9329%, 29.8737%, 42.6246%, Marco 49.52706%	78.8248588754%, 60.1920189345%, 39.2998882688%, 40.2206700042%, 56.5579220767%, Marco 55.0190716319%

- 使用3天播放记录

方法	准确率L1-L5， Macro	召回率L1-L5， Macro	F1值L1-L5， Macro
decision_tree	93.5819%, 89.1506%, 81.7445%, 82.6769%, 85.4587%, Marco 86.52252%	96.0673%, 87.9898%, 77.7835%, 76.8692%, 76.1275%, Marco 82.96746%	94.808314107%, 88.5663966422%, 79.7148251812%, 79.6673458202%, 80.5236732376%, Marco 84.6561109976%
gradient_boosting	94.1348%, 91.7918%, 86.4881%, 88.0939%, 90.8951%, Marco 90.28074%	97.6159%, 89.8798%, 81.0552%, 79.7627%, 80.2306%, Marco 85.70884%	95.8437515307%, 90.8257385925%, 83.683564107%, 83.721549436%, 85.2305458509%, Marco 87.8610299034%
naive_bayes	51.045%, 12.1345%, 24.3346%, 53.8953%, 74.526%, Marco 43.18708%	99.8146%, 0.244%, 1.102%, 7.2038%, 30.6545%, Marco 27.80378%	67.5467289718%, 0.478380740801%, 2.10851522609%, 12.7088929997%, 43.4406998826%, Marco 25.2566435642%
random_forests	93.7153%, 91.7706%, 86.6148%, 89.3994%, 95.1518%, Marco 91.33038%	97.8985%, 89.5398%, 80.5834%, 78.3885%, 76.5344%, Marco 84.58892%	95.7612374166%, 90.6414763839%, 83.4903135838%, 83.5326607807%, 84.8336782097%, Marco 87.6518732749%
xgboost	94.0812%, 91.8408%, 86.8904%, 89.5751%, 95.8437%, Marco 91.64624%	97.7438%, 90.0375%, 81.4891%, 79.353%, 78.196%, Marco 85.36388%	95.877534175%, 90.930210256%, 84.10311819%, 84.1547724778%, 86.1251078369%, Marco 88.2381485871%
svm_linear	69.0313%, 70.769%, 57.9283%, 23.43%, 98.8997%, Marco 64.01166%	92.8141%, 35.337%, 2.5037%, 83.3049%, 51.8142%, Marco 53.15478%	79.1752868025%, 47.1370922097%, 4.79994323239%, 36.573488278%, 68.0018078723%, Marco 47.137523679%
knn	73.8197%, 63.1788%, 54.9371%, 64.6367%, 86.2841%, Marco 68.57128%	86.4733%, 60.0825%, 31.2704%, 31.4356%, 41.8108%, Marco 50.21452%	79.6470596222%, 61.5917607716%, 39.8551191449%, 42.2992568414%, 56.3271019889%, Marco 55.9440596738%

- 使用5天播放记录

方法	准确率L1-L5， Macro	召回率L1-L5， Macro	F1值L1-L5， Macro
decision_tree	94.9904%, 91.8256%, 85.9313%, 86.4954%, 89.304%, Marco 89.70934%	96.7844%, 90.5049%, 84.1409%, 83.0403%, 82.6721%, Marco 87.42852%	95.8790088141%, 91.1604667945%, 85.0266759667%, 84.7326429138%, 85.8601772967%, Marco 88.5317943572%
gradient_boosting	95.4023%, 94.0105%, 89.5471%, 90.5841%, 92.6036%, Marco 92.42952%	98.1599%, 91.9481%, 86.4139%, 84.9863%, 84.9101%, Marco 89.28366%	96.761456811%, 92.9678633314%, 87.9526047782%, 87.6959612535%, 88.5901306362%, Marco 90.7936033621%
naive_bayes	51.0512%, 12.1641%, 24.5965%, 53.4709%, 75.8509%, Marco 43.42672%	99.8146%, 0.244%, 1.102%, 7.2977%, 31.7396%, Marco 28.03958%	67.5521570498%, 0.478403687914%, 2.1094883359%, 12.842638696%, 44.752598522%, Marco 25.5470572583%
random_forests	95.1198%, 94.0339%, 89.9052%, 92.006%, 95.9103%, Marco 93.39504%	98.3789%, 91.807%, 86.1866%, 84.2864%, 82.706%, Marco 88.67298%	96.7219034776%, 92.9071077174%, 88.0066364285%, 87.9771847045%, 88.8200827338%, Marco 90.8865830124%
xgboost	95.4489%, 94.0656%, 90.1292%, 91.8533%, 96.6627%, Marco 93.63194%	98.2666%, 92.2675%, 86.7273%, 85.0717%, 83.4859%, Marco 89.1638%	96.8372574909%, 93.1578742371%, 88.3955315995%, 88.3325293951%, 89.5923976754%, Marco 91.2631180796%
svm_linear	56.1024%, 57.4057%, 55.8351%, 81.2941%, 82.832%, Marco 66.69386%	99.1994%, 10.3936%, 17.1195%, 23.5917%, 76.5683%, Marco 45.3745%	71.6710871163%, 17.6005322627%, 26.2044886669%, 36.5705561472%, 79.577082673%, Marco 46.3247493732%
knn	74.262%, 63.1506%, 55.5335%, 66.0733%, 87.6131%, Marco 69.3265%	86.3104%, 60.8575%, 31.8318%, 31.6832%, 42.6924%, Marco 50.67506%	79.8341797818%, 61.9828485317%, 40.4675830175%, 42.8291434035%, 57.4099099492%, Marco 56.5047329367%

- 使用7天播放记录

方法	准确率L1-L5， Macro	召回率L1-L5， Macro	F1值L1-L5， Macro
decision_tree	95.8911%, 93.4932%, 88.098%, 89.9433%, 90.3949%, Marco 91.5641%	97.3702%, 92.1264%, 88.0222%, 85.3448%, 86.1648%, Marco 89.80568%	96.6249899511%, 92.8047678206%, 88.0600836883%, 87.5837315806%, 88.2291766413%, Marco 90.6605499364%
gradient_boosting	96.2738%, 95.2989%, 91.711%, 92.3918%, 93.8857%, Marco 93.91224%	98.4617%, 93.5089%, 89.6201%, 87.9993%, 87.9959%, Marco 91.51718%	97.3554592096%, 94.3954149162%, 90.6534950828%, 90.142071596%, 90.8454364667%, Marco 92.6783754543%
naive_bayes	51.0548%, 12.188%, 24.5965%, 52.4848%, 78.459%, Marco 43.75662%	99.8146%, 0.244%, 1.102%, 7.3916%, 32.1126%, Marco 28.13296%	67.5553086322%, 0.478422136422%, 2.1094883359%, 12.9582489154%, 45.5726874423%, Marco 25.7348310925%
random_forests	96.0079%, 95.4904%, 92.0266%, 93.7791%, 97.0992%, Marco 94.88064%	98.76%, 93.3404%, 89.5134%, 87.6238%, 86.2665%, Marco 91.10082%	97.3645062046%, 94.4031602065%, 90.7526039048%, 90.5970202525%, 91.3628681569%, Marco 92.8960317451%

xgboost	97.7498%, Marco 95.0194%	88.2298%, 86.9108%, Marco 91.43984%	90.8347971446%, 92.0121922905%, Marco 93.1428585293%
svm_linear	64.0334%, 69.0087%, 50.8359%, 81.7169%, 82.7276%, Marco 69.6645%	96.0666%, 29.1781%, 39.6873%, 27.5435%, 81.0444%, Marco 54.70398%	76.8453594558%, 41.0145304556%, 44.5750838254%, 41.2000950967%, 81.877350285%, Marco 57.1024838237%
knn	74.7694%, 63.1165%, 55.6113%, 66.3236%, 88.1242%, Marco 69.589%	85.9648%, 61.7393%, 32.4241%, 31.9051%, 44.2862%, Marco 51.2639%	79.9772110369%, 62.4203045185%, 40.9641201683%, 43.0843753477%, 58.9483295276%, Marco 57.0788681198%
ml	95.4421%, 94.251%, 90.1917%, 93.7159%, 95.0518%, Marco 93.7305%	98.7142%, 92.6183%, 85.6941%, 82.1014%, 84.0285%, Marco 88.6313%	97.0505777852%, 93.4275174499%, 87.8853956257%, 87.5250227624%, 89.2008800108%, Marco 91.0178787268%

- 使用9天播放记录

方法	准确率L1-L5，Macro	召回率L1-L5，Macro	F1值L1-L5，Macro
decision_tree	96.5758%, 94.4083%, 90.747%, 91.6777%, 93.4074%, Marco 93.36324%	97.6795%, 93.5804%, 89.9094%, 88.9467%, 88.4028%, Marco 91.70376%	97.1245145548%, 93.9925269691%, 90.3262582649%, 90.2915539494%, 90.836220418%, Marco 92.5142148312%
gradient_boosting	96.9325%, 96.2139%, 93.4618%, 93.5726%, 93.9951%, Marco 94.83518%	98.7573%, 94.6396%, 91.5659%, 90.5855%, 90.234%, Marco 93.15646%	97.8363919044%, 95.420257008%, 92.5041367603%, 92.0548241679%, 92.0761579294%, Marco 93.978353554%
naive_bayes	51.0569%, 12.182%, 24.6723%, 50.522%, 82.8397%, Marco 44.25458%	99.8146%, 0.244%, 1.102%, 7.4343%, 32.2482%, Marco 28.16862%	67.5571469859%, 0.478417511669%, 2.10976628657%, 12.9613417213%, 46.424189051%, Marco 25.9061723113%
random_forests	96.7477%, 96.3951%, 93.9053%, 95.2415%, 97.1779%, Marco 95.8935%	98.992%, 94.6317%, 91.745%, 90.3721%, 88.7419%, Marco 92.89654%	97.8569837228%, 95.5052608814%, 92.8125809492%, 92.7429279121%, 92.7685107666%, Marco 94.3372528464%
xgboost	96.9265%, 96.2984%, 93.8677%, 95.088%, 97.8526%, Marco 96.00664%	98.8586%, 94.8218%, 91.9895%, 90.8757%, 89.6236%, Marco 93.23384%	97.8830165615%, 95.5543958736%, 92.9191098236%, 92.934143186%, 93.5574998998%, Marco 94.5696330689%
svm_linear	74.4994%, 65.4175%, 42.9243%, 85.3798%, 78.4181%, Marco 69.32782%	87.8925%, 54.7289%, 42.7558%, 14.0065%, 86.7413%, Marco 57.225%	80.6436591296%, 59.5977543355%, 42.8398843125%, 24.0651310835%, 82.3699763686%, Marco 57.9032810459%
knn	75.1983%, 63.066%, 55.2841%, 65.8764%, 89.7681%, Marco 69.83858%	85.7205%, 62.387%, 32.7341%, 32.016%, 44.6253%, Marco 51.49658%	80.1153858362%, 62.7246624951%, 41.1204786694%, 43.0901443299%, 59.6149571769%, Marco 57.3331257015%

←返回上级 ←返回首页