* **训练数据**

**[地址]**

hdfs://10.51.147.82:9000/user/offline\_feature\_extract/jianyezhou/tuwen\_dwell\_record

用户名：export HADOOP\_USER\_NAME=offline\_feature\_extract

数据量说明：该目录下有三天的图文模型训练样本，但不完整，可以认为被随机采样过了，量约为实际量的1/20

**[数据格式]**

label \t 时间 \t 时长 \t feature1 \t feature1\_value \t feature2 \t feature2\_value

其中：

1. feature\_value目前填充的是我们这边DNN里该特征对应的槽位，可不管，使用时置为１即可
2. feature是一个uint64的整数，64的前16位用于标识该特征所属的特征类，后48位为该特征类计算出的特征值，注意同一条样本在同一个特征类下有可能有多个不同特征值，比如新特tag特征，由于一个新闻有多个tag，一个样本里就会出现多个同类特征。特征前16位对应的整 数与特征名的对应表如下所示

|  |  |
| --- | --- |
| 52173 | user3dayChannelIdCTR |
| 10628 | userLastRefreshTagAvgShowCTR |
| 58987 | userLastRefreshTagIdCTR |
| 61704 | userLastRefreshTagAvgShowPercent |
| 5802 | userLastThreeRefreshTagMaxShowCTR |
| 51730 | userThreeDayGidShowPercent |
| 24941 | userThreeDayGidShowCTR |
| 2362 | user3dayTwoLevelChannelIdCTR |
| 27343 | userTodayChannelIdCTR |
| 48508 | user3dayTopic500IdCTR |
| 59946 | userThreeDayMediaShowCTR |
| 65326 | pVSimScore2 |
| 28533 | userSessionTopic500ShowPercent |
| 45861 | userLastRefreshTopicShowPercent |
| 45620 | userTodayTwoLevelChannelIdCTR |
| 7414 | userTodayTagMaxShowCTR |
| 13233 | userYestodayTagMaxShowCTR |
| 65033 | userSessionTwoLevelChannelShowPercent |
| 42640 | userTodayTagIdCTR |
| 18880 | userLastRefreshTwoLevelChannelShowPercent |
| 63675 | userSessionTwoLevelChannelShowCTR |
| 24346 | userThreeDayEidShowCTR |
| 34935 | userSessionTopic500ShowCTR |
| 38162 | userThreeDayEidShowPercent |
| 47014 | userLastRefreshChannelShowPercent |
| 28788 | userLastRefreshTopic500ShowCTR |
| 30538 | userBeforeYestodayTagAvgShowPercent |
| 53442 | userTodayTagAvgShowCTR |
| 19984 | userSessionTwoLevelChannelIdCTR |
| 55817 | sessionTopic2000ScoreAndCTR |
| 58379 | dssmScore2 |
| 56523 | userSessionTopic500IdCTR |
| 9040 | user3dayTagIdCTR |
| 61805 | readTimeTopic500Hour |
| 58733 | timeCurrent |
| 65514 | readTimeNewsToday |
| 57843 | sessionChannelClickEntropy |
| 56327 | channelFstTopicMax |
| 58889 | crossSessionNumAndTopic500 |
| 54442 | channelFstLevel |
| 59433 | crossNewsCtrAndClick |
| 62992 | TimeSlotCrossTitleTag |
| 61086 | todayTopicShowEntropy |
| 63520 | crossTodayTopic500ClickEntropyAndTopic500Score |
| 54555 | UserLastestTopTxt5StepClickChannelAndDocChannel |
| 64260 | crossSessionNumAndTopicScore |
| 63122 | newsidMd5 |
| 63346 | cross2LVChannelModelAndTopic |
| 56917 | crossSessionChannelClickEntropyAndTopicScore |
| 54244 | crossTitleWordNumAndIsImageSet |
| 50255 | newsTopic500Score |
| 18364 | user3dayTagAvgShowCTR |
| 57568 | userYestodayChannelShowPercent |
| 57341 | crossIRCMJaccard3ScoreAndTopic500Score |
| 51259 | crossTopic500AndDssmScore |
| 52146 | newsTagSumScoreNew |
| 57246 | todayChannelClickEntropy |
| 55073 | crossTopic500AndHQTagScore |
| 7341 | userThreeDayKeyWordIdCTR |
| 26303 | newsSecondChannel |
| 53841 | crossSessionTopicClickEntropyAndTopicScore |
| 9888 | crossUserThreeDaysCTR\_tag |
| 23345 | userLastRefreshTwoLevelChannelIdCTR |
| 54269 | newsMediaMd5 |
| 64859 | userTodayTopicShowCTR |
| 53566 | crossChannelAndTimeliness |
| 50450 | crossUserLocationAndLocation |
| 55246 | UserLastestTopTxt1StepClickTopic500AndDocTopic500 |
| 50230 | userBeforeYestodayTwoLevelChannelShowCTR |
| 25425 | crossSessionDwellTimeAndTopicScore |
| 39753 | crossTopicAndTagScore |
| 56051 | CrossUserGroupIdAndMaxTopic2000 |
| 4639 | readLenTopic500ThreeDay |
| 36280 | readLenTopic2000Today |
| 56612 | newsTopic2000Score |
| 7648 | user3dayTopicIdCTR |
| 26958 | nextReadScore |
| 56099 | user3dayTopic500ShowCTR |
| 20539 | CrossUserGroupIdAndMaxTopic130 |
| 51484 | UserLastestTopTxt5StepClickTopic500AndDocTopic500 |
| 11383 | newsIsListImage |
| 52970 | userBeforeYestodayTagMaxShowCTR |
| 35142 | titleImageSetNum |
| 38608 | userYestodayTopic500ShowPercent |
| 6533 | crossSessionNumAndTopic |
| 38249 | crossTodayChannelClickEntropyAndTopicScore |
| 17093 | readLenTopic2000ThreeDay |
| 16237 | cross2LVChannelAndDssmScore |
| 15327 | userContextArea |
| 53435 | sessionTopic500ScoreAndUserTopic |
| 50932 | crossTopicModelAndMedia |
| 21145 | readTimeNewsThreeDay |
| 20424 | cross2LVChannelAndHQTagScore |
| 18235 | UserLastestTopTxt5StepClickSecChannelAndDocSecChannel |
| 56337 | crossTopic500ModelAndMedia |
| 40233 | cross2LVChannelModelAndTopic500 |
| 60733 | crossNewsidAndUserLocation |
| 38172 | newsTitleRow |
| 22164 | threeDayTopicShowEntropy |
| 52871 | cross3dayDwellTimeAndTopicScore |
| 14542 | userLastRefreshChannelIdCTR |
| 14865 | newsQualityFrmOmg |
| 14889 | todayChannelShowEntropy |
| 25136 | sessionTopicScoreAndUserTopic |
| 61142 | crossTagBetaModelAndMedia |
| 20936 | userTodayChannelShowCTR |
| 57234 | userEnvironmentSexAgeOSBRANDNetWorkMedia |
| 57810 | userEnvironmentSexAgeOSBRANDNetWorkTopic500 |
| 18921 | channelFstTopicScore |
| 17154 | TimeSlotCrossTopic2000 |
| 34042 | userYestodayTopicShowPercent |
| 34332 | TimeSlotCrossChannel |
| 27826 | UserLastestTopTxt10StepClickTopic500AndDocTopic500 |
| 5017 | crossTopic500ModelAndTopic |
| 31054 | crossCtrOfNewsAndNewsTagSumScore |
| 22414 | crossNewsIdAndTagScore |
| 11492 | ctrOfMedia |
| 52593 | crossCtrOfNewsAndNewsTopic500ScoreV2 |
| 12533 | readLenTopic500Today |
| 46402 | userBeforeYestodayTopic500ShowPercent |
| 21031 | userTodayTopicIdCTR |
| 38127 | crossUserThreeDaysCTR\_topic500 |
| 29015 | userBeforeYestodayTopic500ShowCTR |
| 6517 | userYestodayChannelShowCTR |
| 26846 | crossCtrOfNewsAndNewsTopic2000Score |
| 25070 | crossSessionDwellTimeAndDssmScore |
| 6956 | userRecent3SeqTitleTagCTR |
| 22699 | crossSourceCtrAndClick |
| 20988 | user3dayTwoLevelChannelShowCTR |
| 24479 | newsFreshLevel |
| 52106 | sessionTopic2000ScoreAndUserTopic |
| 54733 | cross2LVChannelModelAndTag |
| 25139 | crossUserLocationAndMedia |
| 14319 | userThreeDayTermMaxShowCTR |
| 54086 | UserLastestTopTxt1StepClickChannelAndDocChannel |
| 63435 | NewsHQTitleTagEachScore |
| 6165 | crossMediaIdAndDssmScore |
| 44381 | crossNewsIdAndTopic500Score |
| 14110 | userCTR |
| 18346 | newsSource |
| 10599 | crossLastRefreshDwellTimeAndDssmScore |
| 43673 | crossSessionNumAndTopic500Score |
| 2961 | sessionChannelShowEntropy |
| 4747 | timeLiness |
| 10722 | crossTopicModelAndTopic |
| 3276 | userTodayCTR |
| 17100 | crossCtrOfNewsAndDssmScore |
| 44429 | crossHQTagScoreAndDssmScore |
| 64461 | channelFstLevelScore |
| 47225 | crossTopic500ModelAndTag |
| 25377 | crossNewsidAndUserAge |
| 2137 | readTimeTopic500Today |
| 14078 | sessionTopic500ScoreAndNewsTopic |
| 23061 | userTodayTagAvgShowPercent |
| 10791 | videoUserLastRefreshNSecondChannelShowPTR |
| 15451 | cross2LVChannelModelAndChannel |
| 19175 | userLastRefreshTagMaxShowCTR |
| 4247 | cross3dayDwellTimeAndDssmScore |
| 47542 | channelScdLevelScore |
| 60348 | userThreeDayKeyWordAvgShowPercent |
| 24181 | userTodayTopicShowPercent |
| 44766 | crossNewsIdAndTopicScore |
| 64550 | userSessionTagIdCTR |
| 54326 | userSessionTagAvgShowPercent |
| 3405 | readTimeTopic2000Hour |
| 12416 | userSex |
| 45450 | newsIsBigImage |
| 17335 | userYestodayTopic500ShowCTR |
| 10132 | userTodayChannelShowPercent |
| 4724 | crossTopic500ModelAndTopic500 |
| 4632 | crossNewsIdAndHQTagScore |
| 19842 | recallFrom |
| 19236 | crossNewsTitleTagAndUserTopChannel |
| 21482 | bodyWordNum |
| 2772 | UserLastestTopTxt10StepClickSecChannelAndDocSecChannel |
| 39220 | crossTopicAndHQTagScore |
| 62787 | pVSimScore |
| 13862 | crossSessionTopic500ClickEntropyAndTopic500Score |
| 3620 | cross2LVChannelModelAndMedia |
| 60357 | useridMd5 |
| 52064 | crossTodayDwellTimeAndTopicScore |
| 65261 | UserTopVideoTagNewsFreshLevel |
| 1524 | crossTopicScoreAndHQTagScore |
| 45760 | crossUserSexAgeAndNewCategoryL2 |
| 1155 | newsBigSource |
| 20932 | crossMatchTagBetaEntropyAndTagBetaScore |
| 29377 | showListLoc |
| 36279 | userAge |
| 9783 | titleWordNum |
| 6890 | userTodayTwoLevelChannelShowPercent |
| 18840 | user3dayTagAvgShowPercent |
| 42008 | crossTodayDwellTimeAndDssmScore |
| 10286 | crossLastRefreshChannelClickEntropyAndTopicScore |
| 11238 | readLenNewsThreeDay |
| 41126 | userLastSixRefreshTagIdCTR |
| 21303 | crossNewsidAndUserTopChannel |
| 47338 | cross3DayTopicClickEntropyAndTopicScore |
| 3569 | user3dayTopicShowPercent |
| 9215 | sessionTopicShowEntropy |
| 15252 | mediaScore |
| 60413 | userSessionTagAvgClickPercent |
| 14020 | user3dayTagAvgClickPercent |
| 32937 | crossLastRefreshTopicClickEntropyAndTopicScore |
| 19464 | onlineScore |
| 6965 | sessionTopic500ScoreAndRank |
| 41685 | userYestodayTwoLevelChannelShowCTR |
| 11162 | TimeSlotCrossSubChannel |
| 11214 | crossCtrOfNewsAndNewsTopic2000ScoreV2 |
| 56648 | userYestodayTagAvgShowPercent |
| 25793 | cross3DayChannelClickEntropyAndTopicScore |
| 56152 | crossAgeChannel |
| 32911 | crossLastRefreshDwellTimeAndTopicScore |
| 38066 | TitleTagEachScoreFix |
| 28250 | TimeSlotCrossRecallFrom |
| 12414 | crossTopic500AndTagScore |
| 59858 | crossTopicModelAndTopic500 |
| 10566 | userYestodayTopicShowCTR |
| 56953 | userSessionTagMaxShowCTR |
| 63539 | crossNewsIdAndDssmScore |
| 27629 | readLenNewsToday |
| 4820 | userBeforeYestodayTagIdCTR |
| 10431 | mediaScore2 |
| 8273 | userEnvironmentCityAndMedia |
| 43664 | crossBodyWordNumAndImageNum |
| 56677 | userLastRefreshTopic500ShowPercent |
| 18622 | user3dayChannelShowPercent |
| 31814 | readTimeNewsHour |
| 31615 | titleTagSumScoreNew |
| 27972 | crossUserLocationAnd2LVChannel |
| 28398 | cross2LVChannelAndTagScore |
| 11551 | userLastSixRefreshTagMaxShowCTR |
| 64028 | titleHQTagSumScoreNew |
| 35197 | userEnvironmentSexAgeOSBRANDNetWorkTopic |
| 32644 | titleIsImageSet |
| 32463 | userTodayTagAvgClickPercent |
| 27934 | crossChannelModelAndChannel |
| 41002 | crossAreaNewsArea |
| 30055 | lastRefreshChannelShowEntropy |
| 39030 | crossMatchTagBetaEntropyAndHQTagScore |
| 57435 | crossUserSexAgeAndTopic |
| 16036 | userYestodayTagIdCTR |
| 32805 | userLastRefreshTopicIdCTR |
| 9232 | crossTopicModelAndTag |
| 18075 | newsTopic2000\_refine\_cross\_userTopChannel |
| 29274 | crossCtrOfNewsAndContentUserTopicScore |
| 61844 | crossSexChannel |
| 13441 | userSessionTagAvgShowCTR |
| 57392 | userLastRefreshTopic500IdCTR |
| 65266 | cross2LVChannelModelAnd2LVChannel |
| 34152 | nextReadScore2 |
| 65116 | UserLastestTopTxt1StepClickSecChannelAndDocSecChannel |
| 25995 | readTimeTopic2000Today |
| 5985 | crossLastRefreshTopic500ClickEntropyAndTopic500Score |
| 43589 | threeDayChannelShowEntropy |
| 50651 | UserTopVideoTagNewsChannel |
| 38775 | userEnvironmentSexAgeOSBRANDNetWorkTopic2000 |
| 22345 | userLastRefreshTagAvgClickPercent |
| 32819 | userThreeDayTermAvgShowPercent |
| 33670 | threeDayChannelClickEntropy |
| 55515 | sourceQuality |
| 34206 | crossSessionNumAndTagScore |
| 50122 | user3dayTopic500ShowPercent |
| 34378 | crossCtrOfNewsAndNewsHQTagSumScore |
| 34300 | readLenTopic2000Hour |
| 53114 | crossTitleWordNumAndListImage |
| 36402 | readLenTopic500Hour |
| 34611 | crossSessionNumAndHQTagScore |
| 45712 | userLastestTopTxt10StepClickChannelAndDocChannel |
| 40719 | sessionTopicScoreAndRank |
| 31009 | crossTodayTopicClickEntropyAndTopicScore |
| 60179 | crossMediaIdAndTagScore |
| 63903 | crossUserSexAgeAndMedia |
| 45809 | crossTopicModelAndChannel |
| 1791 | crossNewsMediaAndUserTopChannel |
| 31263 | newsEventLevel |
| 28935 | sessionTopic2000ScoreAndNewsTopic |
| 38690 | ctrOfHourNews |
| 55331 | crossTopicAndDssmScore |
| 38086 | crossMatchTagEntropyAndTagScore |
| 33033 | cross3DayTopic500ClickEntropyAndTopic500Score |
| 17766 | TimeSlotCrossTopic500 |
| 39709 | crossTopic500ScoreAndTagScore |
| 16540 | crossTopicScoreAndTagScore |
| 40382 | newsTopic500Score2 |
| 65210 | crossTopicModelAnd2LVChannel |
| 8951 | userThreeDayMediaIdCTR |
| 45154 | user3dayTagMaxShowCTR |
| 4374 | userLastRefreshTwoLevelChannelShowCTR |
| 38818 | crossNewsidAndUserSexAge |
| 40425 | readTimeTopic500ThreeDay |
| 16396 | userBeforeYestodayTwoLevelChannelShowPercent |
| 41191 | crossIRCMJaccard3ScoreAndDssmScoreTagBetaScore |
| 51408 | crossTagScoreAndDssmScore |
| 30200 | crossCtrOfNewsAndNewsTopic500Score |
| 39302 | crossNewsidAndUserSex |
| 49954 | user3dayTwoLevelChannelShowPercent |
| 43879 | userTodayTwoLevelChannelShowCTR |
| 60210 | crossSessionNumAnd2LVChannel |
| 42411 | crossTopic500ScoreAndHQTagScore |
| 57362 | newsTopic500\_refine\_cross\_userTopChannel |
| 48962 | crossUserSexAgeAndTopic500 |
| 43441 | CrossUserGroupIdAndMaxTopic500 |
| 63329 | readLenNewsHour |
| 48441 | sessionTopicScoreAndNewsTopic |
| 42898 | crossSessionNumAndDssmScore |
| 12341 | userLastRefreshTopicShowCTR |
| 45865 | crossChannelAndArea |
| 45260 | ctrOfNews |
| 37182 | user3dayChannelShowCTR |
| 34034 | crossHourNewsCtrAndClick |
| 41592 | crossSessionNumAnd2LVChannelScore |
| 31564 | crossTitleWordNumAndBigImage |
| 24823 | crossTopic500ModelAnd2LVChannel |
| 23761 | userLastRefreshChannelShowCTR |
| 46737 | entropyOfChannelModel |
| 42544 | userBeforeYestodayChannelShowCTR |
| 47515 | crossSessionNumAndChannel |
| 6088 | crossTagBetaModelAndChannel |
| 58061 | userYestodayTwoLevelChannelShowPercent |
| 50015 | readTimeTopic2000ThreeDay |
| 43409 | user3dayTopicShowCTR |
| 47018 | userBeforeYestodayChannelShowPercent |
| 50787 | crossUserThreeDaysCTR\_topic |
| 48710 | crossCtrOfNewsAndContentUserTopicScoreV2 |
| 47767 | sessionTopic2000ScoreAndRank |
| 39853 | crossTagBetaModelAndTag |
| 32998 | UserTopVideoTagMaxTopic |
| 40655 | crossUserSexAgeAndTitleTag |
| 3599 | userTopChannel |
| 37447 | crossUserSexAgeAndNewTag |
| 48022 | newsTopic\_refine\_cross\_userTopChannel |
| 45817 | crossUserGroupIdAndTags |
| 44134 | userThreeDayTermIdCTR |
| 56144 | crossUserClickIdAndDocId |
| 42033 | newsItem |
| 17883 | crossUserSexAgeAndTitle |
| 54981 | CrossUserTopTxtChannelTagAndTxtDocInfo |
| 32969 | UserTopVideoTagNewsTag |