# 冷热频道区分训练与延时跳数统计分析

## 一、实验环境

Windows10、matlab2016a、LightNet深度学习框架（GUI-version 2）

## 二、实验模型与流程

### 1.基于长短期记忆神经网络(LSTM)推荐法

图2.1为长短期记忆神经网络(LSTM)推荐法流程示意图，本文将以某一IPTV用户A为例子，对长短期记忆神经网络(LSTM)推荐法进行详细阐述：

1）为训练集选取大小，要这一天为用户进行推荐，选取用户A在当天的过去天内，即示意图中这一时间内，用户A的观看数据作为训练数据集。

2）对训练数据进行噪声数据的过滤，如去掉观看时长小于10s等数据预处理，同时，按序，将用户的所有记录序列化，只保留用户每次观看时所观看的频道号，构成的有序频道序列。频道序列指的是，用户A观看完频道后再观看频道。

3）训练需要确定输入序列的长度，本文用S表示序列长度，即构造输入样例的序列长度，如图2.1所示，设序列长度为，第一个输入样例由第一个输入序列以及第一个输入标签构成，[]构成第一个输入序列，[]为第一个输入标签；[]构成第二个序列，[]为第二个序列的标签，以此类推，最终将会得到个输入样例。例如，用户A在所选训练集时间区间内的观看序列为[7，8，9，33，8，9]，假如序列长度为4，那么最终会得到3个序列，[7，8，9，33]、[8，9，33，8]、[9，33，8，9]，且3个序列所对应的标签分别为[33]，[8]，[9]。

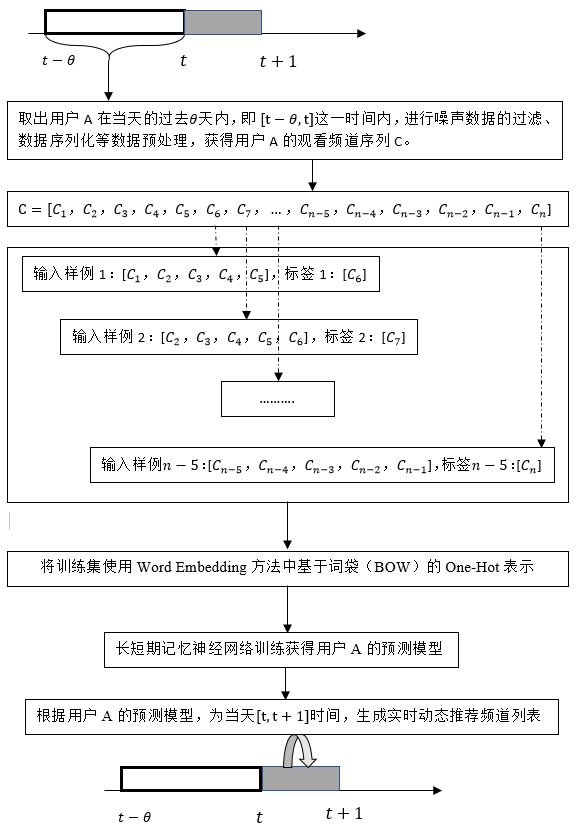


图 2.1长短期记忆神经网络(LSTM)推荐法流程示意图

4）经过前三个步骤对训练集的提取以及处理后，将训练集输入长短期记忆神经网络中的LSTM进行训练，并获得对应的用户A的预测模型.

5）根据用户A的预测模型，为当天[t,t+1]时间内，用户A观看频道的实时观看记录，生成相应实时动态推荐频道列表。

### 2.基于LSTM冷热频道区分训练推荐法

设定冷热频道阈值参数为P，用于划分用户的冷门频道与热门频道集合。将一段时间内用户观看某一频道的次数，占该时间段内用户观看总次数N的百分比称作观看频道频率占比，冷热频道阈值P与频道频率占比的关系表示如下，

式(2-1)

式(2-2)

根据冷热频道阈值参数可将用户观看频道划分为热门频道以及冷门频道，进而获得用户的热门频道观看序列以及冷门频道观看序列，再将此两个序列分别采用长短期记忆神经网络(LSTM)推荐法，以观察推荐效果。

基于LSTM推荐法为热门频道序列进行推荐的流程将以图2.2为例进行描述，该流程主要包含以下步骤：

1）选取用户A的训练集中内，统计每个频道的观看频率，并根据前文中的式(2-1)，分别获得热门频道集合以及冷门频道集合。

2）在用户A的原始观看序列中，根据与，提取热门频道观看序列]以及冷门频道观看序列。

3）将这两个序列分别按照2.1节LSTM推荐法中对输入训练集的数据预处理进行预处理，并使用LSTM进行训练，得到用户的热门频道预测模型与冷门频道预测模型。

4）获得热门频道预测模型以及冷门频道预测模型后，按照2.1中LSTM推荐法中步骤5）分别进行预测，如图2.2中以热门频道预测模型为例，对原始序列中的热门频道序列进行逐一推荐。

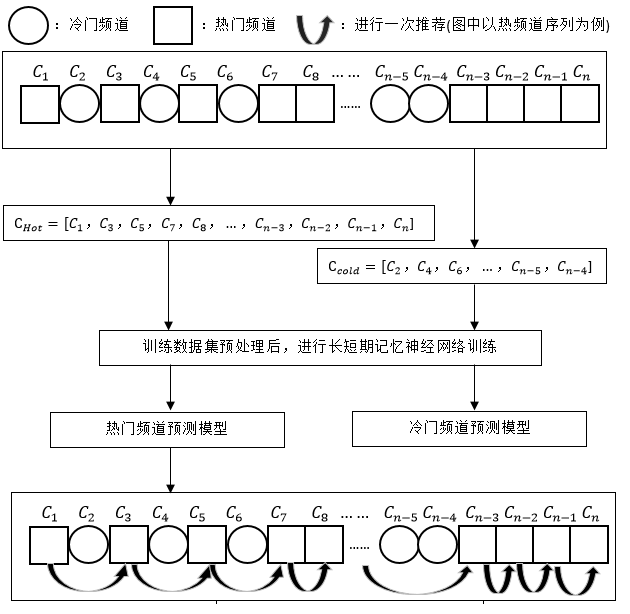


图2.2 基于LSTM的冷热频道区分训练推荐流程示意图

### 3.单一训练、区分训练及延时跳数

单一训练(standalone learning,SL)表示对频道不进行冷热区分进行训练，即将用户滑动窗口内去除噪声后的数据直接进行训练，在预测时也不会对用户的冷热门频道进行区分，即采用2.1节介绍的推荐流程。

区分学习(distinguished learning,DL)则是先设定冷热阈值P，再对冷热频道分别进行训练和预测，即采用2.2节介绍的推荐流程。

延时跳数(DelayStep,DS)，为本文引入的一个参数，用于研究DL方法将提取出来的热频道序列还原到混合频道序列时，相邻两个热频道的距离。设定SL方法延时跳数DS为1

## 三、实验指标与结果

LSTM训练参数配置：训练集大小设置为、输入序列长度设置为

### 1. 推荐准确率与推荐召回率

推荐准确率Acc为测试阶段，用户每进行一次切换，便会开始预测用户下一次可能切换的频道，若预测的频道命中用户切换的频道，则视为推荐成功具体公式可以描述为以下：

式(3-1)

其中表示第d天对用户u的预测准确率，表示向用户u进行推荐的次数，表示向用户u的第i次切换频道时推荐N个频道供用户选择，表示第i次切换时所要的频道，U表示所有用户的集合，为训练集合大小。为了文章更简洁统一，本文用来表示用户u的相对于天数的平均推荐准确率，用表示所有用户集的平均推荐准确率。

推荐召回率Recall，具体公式如下

式(3-2)

其中表示第d天对用户u的预测准确率，表示向用户u进行推荐的次数，表示向用户u的第i次切换频道时推荐N个频道供用户选择，表示第i次切换时所要的热门(冷门)频道，表示参与测试热门(冷门)频道数目，为训练集合大小，为了文章更简洁统一，本文用来表示用户u的相对于天数的平均推荐召回率，用表示所有用户集的平均推荐召回率。

### 2.SL推荐实验结果

样本为201408用户观看记录，最活跃的100位用户的测试结果，如下：

|  |  |  |
| --- | --- | --- |
| **SL** | **推荐准确率** | **推荐召回率** |
| **TOP 1** | 16.01% | 14.22% |
| **TOP 2** | 26.95% | 23.89% |
| **TOP 3** | 35.99% | 31.87% |
| **TOP 4** | 43.29% | 38.32% |
| **TOP 5** | 50.08% | 44.67% |

图3.1 LSTM推荐法实验结果

### 3.DL推荐实验结果

样本为201408用户观看记录，最活跃的50位用户的测试结果，如下：

图3.2 冷热频道阈值与热门频道区分推荐准确率统计图

图3.3 冷热频道阈值与冷门频道区分推荐准确率统计图

图3.4冷热频道阈值与热门频道区分推荐召回率统计图

图3.5冷热频道阈值与冷门频道区分推荐召回率统计图

综合召回率可以看出，当划分阈值在4.5%与4%之间，热门频道推荐召回率可以能达到一个峰值，同时也存在不错的推荐效果。

### 4.DS指标及分析结果

本文将分析DS的平均值以及命中平均值，计算公式如下

其中表示用户u的平均延时跳数，表示向用户u进行推荐的次数，表示向用户u的第i次推荐频道时的延时跳数，为该用户的有效测试天数，SL的DS均为1

统计单个用户推荐命中平均时延（RecommDelayStep，RDS）的公式为（当天推荐未命中视为0）

其中表示用户u的推荐命中平均延时跳数，表示向用户u进行推荐的次数，表示向用户u的第i次推荐频道时的延时跳数，为该用户的有效测试天数，表示向用户u的第i次切换频道时推荐N个频道供用户选择，表示第i次切换时所要的热门(冷门)频道。

实验结果如下：

样本为201408用户观看记录，最活跃的50位用户的测试结果，分别设定冷热频道阈值为4%、4.5%

|  |  |  |  |
| --- | --- | --- | --- |
| DL时延跳数 | | 4%TOP50 | 4.5%TOP50 |
| 热频道平均延时跳数 | | 2.082368129 | 2.162580045 |
| 热频道命中的平均延时跳数 | TOP1 | 1.822737655 | 1.907262787 |
| TOP2 | 1.993005959 | 2.01031667 |
| TOP3 | 1.988233592 | 2.06097981 |
| TOP4 | 2.012395941 | 2.101003884 |
| TOP5 | 2.050779934 | 2.134631271 |
| 冷频道平均延时跳数 | | 2.259644092 | 2.169235678 |
| 冷频道命中的平均延时跳数 | TOP1 | 2.268592109 | 2.176437882 |
| TOP2 | 2.140812376 | 2.105421277 |
| TOP3 | 2.194638763 | 2.137078538 |
| TOP4 | 2.183058782 | 2.163311525 |
| TOP5 | 2.236740664 | 2.184639483 |
| 统计单个用户推荐平均时延的公式为 DelayStep = （当天每次推荐延时跳数总和/当天推荐次数）/有效测试天数 | | | |
|
|
| 统计单个用户推荐命中平均时延的公式为（当天推荐未命中视为0） RecommDelayStep = （当天推荐命中的延时跳数总和/当天推荐命中次数）/有效测试天数 | | | |
|
|

样本为201408用户观看记录，最活跃的300位用户的测试结果，分别设定冷热频道阈值为4%、4.5%