

流媒体点播学习系统自适应性研究*

蒋冬梅

(北京市海淀区职工大学,北京 100083)

摘要: 针对为流媒体学习系统提供个性化点播服务的问题,通过研究学习系统自适应性实施的相关文献,给出一个基于朴素贝叶斯分类的自适应流媒体点播系统的实现模型。为了提高分类诊断性能,适合流媒体学习特性,重点讨论为不同知识点类型赋予不同权重的基于概率加权的贝叶斯分类算法和加入时序的贝叶斯分类算法。实验结果表明,算法设计合理、科学。

关键词: 流媒体点播;自适应性;个性化服务;朴素贝叶斯分类

中图分类号: TP311.1

文献标识码: A

文章编号: 1674-7720(2011)17-0085-03

The adaptability of research in streaming media VOD learning system

Jiang Dongmei

(Beijing Haidian Community College, Beijing 100083, China)

Abstract: Aiming at for streaming media learning system providing personalized VOD service problems, this paper presents an adaptive streaming media on demand system implementation model based on Naive Bayesian Classifier by researching on the related literature of the adaptability implementation in learning system. In order to improve the classification performance, suitable for streaming media studying characteristics, this paper focuses on two methods setting different weights of different knowledge based on probability inference for using with Naive Bayesian Classification and Naive Bayesian Classification by adding timing. Experimental results indicate that the methods are reasonable, scientific.

Key words: streaming media demand; adaptability; personalized service; Naive Bayesian Classification

随着基于 Web 的现代远程教育的发展,模拟教师的智能,按学习者的学习行为和学习状态动态导航学习内容,提供个性化教学服务,已成为开发远程教育系统的需求。

目前,国内外研究人员已开展了学习系统的自适应性研究,并取得了一定的成果。系统自适应性的关键是实现对学习者的学习情况的诊断和资源推荐,通常采用分析学习者的学习特征与资源特征间关系的方式。Brusilovsky^[1]开发的自适应学习系统,根据学习者的学习能力进行学习情况诊断和资源推荐。CHEN C M 等^[2]开发的自适应学习系统,根据学习者能力和学习资料的难度推荐资源。WANG H C 等^[3]开发的自适应学习系统,根据学习者特征和学习资源特征的最大相似度匹配策略进行学习者情况的诊断和资源推荐。知识是分层次的,以知识点来推进整个学习过程,更符合认知规律。但

上述系统未考虑某一知识点对整个单元知识体系的影响,忽略了从流媒体对象的角度考察学习者交互式请求的学习特征,不能充分反映学习者的学习适应性。此外,由客观题测试得到的测试结果并不能充分反映学习者的学习情况。因此,本文从学习者学习特征、知识点类别和测试结果出发,预测学习者的学习情况和知识水平,并根据学习者的学习情况与学习者的知识水平最大匹配策略推荐学习资源。

1 学习者流媒体点播自适应性诊断

对于学习者学习某一单元的流媒体课件后是否能进入下一单元的预测属于数据挖掘和人工智能的范畴。

1.1 自适应点播实现过程

自适应点播系统模型如图 1 所示。系统定期或触发式地检查服务器日志,从服务器向用户流化数据时产生的流化日志条目和用户流化到缓存的数据进行访问时产生的播放日志条目中了解用户访问流媒体课件的

* 基金项目:北京广播电视大学基层电大科研基金项目(07JCSJ05)

情况,如用户的IP地址、访问ID、访问的视频名称、起始时间、观看视频的时间以及前跳、暂停、后跳交互请求等^[4]。经过一个单元的学习后,得到一张学习过程记录表,其形式如表1所示。系统根据学习特征自动诊断学习者学习情况,即内容容易、内容较难、内容难。

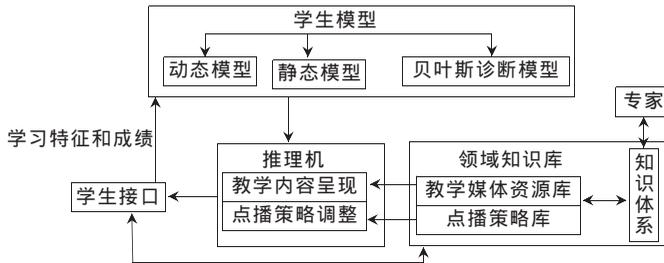


图1 自适应点播系统模型图

表1 学习过程记录表

学生 (ID)	课程 (KC)	单元 (DY)	学习特征(TZ)		
			快进	暂停	快退
01	网络	IP地址	1次	3次	0次
01	网络	IP报文格式	4次	10次	6次
.....

学习者学习一个单元后进入单元测试,并自动生成单元成绩测试表,如表2所示。系统根据各类知识点成绩预测学习者拥有该单元的知识水平,即水平高、水平一般、水平较低、水平极低。

表2 单元成绩表

课程(KC)	单元(DY)	知识点(ZSD)	知识点类别(LB)	成绩(CJ)
网络	IP地址	分类IP	记忆	8
网络	IP地址	地址划分	过程	6
.....

系统自适应地设置视频点播导航及相关资源的调整策略如下:

R1:IF学习者感觉学习内容难 and(学习者当前的知识水平极低 or 低)THEN 点播导航为前驱视频单元及提供相应学习资源链接。

R2:IF学习者感觉学习内容难 and(学习者当前的知识水平一般 or 高)THEN 导航为当前单元练习测试及提供相应学习资源链接。

R3:IF(学习者感觉学习内容较难 or 容易)and(学习者当前的知识水平极低)THEN 点播导航为当前视频单元及提供相应学习资源链接。

R4:IF(学习者感觉学习内容较难 or 容易)and(学习者当前的知识水平较低)THEN 导航为当前单元练习测试及提供相应学习资源链接。

R5:IF(学习者感觉学习内容较难 or 容易)and(学习者当前的知识水平一般 or 高)THEN 点播导航为后继视频单元及提供相应学习资源链接。

1.2 朴素贝叶斯分类器

对学习者的学习情况和知识水平分类诊断需要进行

数据挖掘。常用的分类预测方法很多,贝叶斯分类法因简单易行、分类效果较好而被广泛应用。

朴素贝叶斯分类基于类条件独立性假设,在已知先验概率与条件概率情况下进行分类识别。令 $U=\{X_1, X_2, \dots, X_n, C\}$ 是有 n 个属性的离散型随机变量的有限集,具有 m 个类的类集 $C=\{C_i, i=1, 2, \dots, m\}$, 类 C_i 发生的概率 $P(C_i)$ 是先验概率,由训练集数据计算所得。由于拾取了观测数据 X ,对分类结果有了新的认识,根据式(1)找到具有最大概率 $P(C_j|X)$ 的类作为分类预测的结果,这就是贝叶斯估计。

$$\exists i, P(C_i|X) = \max_j (P(C_j|X)) = \max_j P(C_j) \prod_{k=1}^n P(X_k|C_j) \quad (1)$$

其中, $P(C_j) = S_j/S$ 为类 C_j 的训练集样本数/训练集样本总数; $P(X_k|C_j) = S_{jk}/S_j$ 为属于 C_j 类且属性 A_k 取值 a 的训练样本数/属于类 C_j 的训练样本数。

1.3 知识水平分类诊断

由于某单元各类知识点的分数段与预测类别之间存在或大或小的关联,需尝试对属性与类别之间的关联进行量化,放松朴素贝叶斯中条件独立假设,以提高贝叶斯分类器的准确度和分类效率。特征加权的贝叶斯分类的模型如式(2)、(3)所示。

$$\exists i, P(C_i|X) = \max_j (P(C_j|X)) = \max_j P(C_j) \prod_{k=1}^n P(X_k|C_j)^{w_k} = \max_j (P(C_j) \prod_{k=1}^n (\frac{S_{jk}}{S_j})^{w_k}), 1 \leq i, j \leq m, i \neq j \quad (2)$$

$$\exists i, P(C_i|X) = \max_j (P(C_j|X)) = \max_j P(C_j) \prod_{k=1}^n W_{kj} P(X_k|C_j) = \max_j P(C_j) \prod_{k=1}^n W_{kj} \times \frac{S_{jk}}{S_j}, 1 \leq i, j \leq m, i \neq j \quad (3)$$

参考文献[5]提出了通过有序加权算子作为概率乘积的权重办法,参考文献[6]提出用ID3信息增益的决策树技术确定权重,参考文献[5]和参考文献[6]忽略了特征属性与其他各个类别间的相关度,并需要考虑权值为0的问题。参考文献[7]用互信息可信度的方式标识条件属性与类别之间相互关联的强弱程度,进一步提高分类准确率,但它的计算时间长。为此,将特征权值改进为式(4):

$$W_{kj} = \frac{m + P(C_j \wedge A_k = a)}{m} = \frac{m + \frac{S_{jk}}{S'_k}}{m}, 1 \leq j \leq m, 1 \leq k \leq n \quad (4)$$

其中, S'_k 为训练集中属性 A_k 取值 a 的样本数。

由于 $W_{kj} \geq 1$,所以用式(3)比式(2)更能突显重要属性对分类的影响。基于概率加权的朴素贝叶斯分类算法如下:

(1)数据预处理,从成绩表中选取训练样本和测试样本。

(2)扫描训练集生成统计表,统计具备某种知识水平时各类知识点的不同分数段的样本数,以及各类知识点

的不同分数段的样本数。

(3) 概率参数学习,生成概率表。

(4) 权值参数学习,计算属性 X_k 的权 W_{kj} 。

(5) 生成加权贝叶斯分类器,按式(3)生成概率表及权值列表。

(6) 调用概率表及权值列表得分类结果。

1.4 学习情况分类诊断

根据学习者操作媒体的特征预测学习者的学习情况是容易、较难或难,可使用朴素贝叶斯分类的方法实现。贝叶斯分类是在某一指定的时刻进行的。在流媒体课件的播放过程中,学习者对媒体的操作是不断变化的,它与时间 t 有关,因此,学习者的学习情况集要用时序状态集来表示,即 $C=\{C_i, i=1, \dots, m(m=3)\}, t=\{\text{容易, 较难, 难}\}, t=1, 2, 3, \dots, T$ 。学习情况 C_i 的概率 $P_t(C_i)$ 表示它们的统计特征。这里, t 为广义时间单位,即定期诊断或检测次数乘以间隔时间得到的实时时间^[7]。相应地,任何时刻学习情况 C_i 的后验概率公式为:

$$P_i(C_i|X_j) = \frac{P(X_j|C_i)P_i(C_i)}{P(X_j)} = \frac{P(X_j|C_i)P_i(C_i)}{\sum_{i=1}^n P(X_j|C_i)P_i(C_i)} \quad (5)$$

加入时序 t 的朴素贝叶斯分类过程如下:

(1) 初始化先验信息: $t=0$, 根据历史数据计算 $P_0(C_i)$ ($i=1, 2, \dots, m$)。

(2) 以 $P_t(C_i)$ 为先验概率,按式(5)计算 $t=t+1$ 时刻的后验概率 $P_t(C_i|X_j)$ 。

(3) 若 $t \leq T$, 则返回(2), 否则输出预测结果。

2 实验结果及分析

为了验证算法的有效性,基于校园局域网环境,对23名学习者学习网络基础、C# 程序设计两门课程流媒体课件的部分数据,用朴素贝叶斯分类法 NB、概率加权的贝叶斯分类法 PWNB、加入时序的贝叶斯分类法 TNB 进行测试,结果如表3所示。结果表明,上述贝叶斯分类器预测学习者的学习情况和知识水平的算法简单有效,根据学习者的学习情况与学习者的知识水平最大匹配策略推荐资源的方法可行。

以上讨论了学习系统中自适应流媒体点播系统的实现过程及学习者学习情况、知识水平的诊断流程,设计了一个自适应流媒体点播系统模型,重点讨论了贝叶

表3 实验结果表

数据集	成绩库 样本数	学习特征 初始样本数	准确率			
			NB	PWNB	TNB	
网络	IP 地址	69	138	84.2	87.1	85.6
	IP 报文	23	69	85.6	82.3	80.7
C#	If 语句	37	69	86.8	83.6	83.6
	实例应用	31	69	80.6	78.7	80.7

斯分类算法,并应用于实际教学。通过学习者学习过程中反馈的数据,评估系统及算法设计的合理性、科学性。

参考文献

- [1] BRUSILOVSKY P. Layered evaluation of adaptive learning systems [J]. Int. J. Cont. Engineering Education and Life Long Learning, 2004, 14(4):402-421.
- [2] CHEN C M, LEE H M, CHEN Y H. Personalized E-learning System Using Item Response Theory [J]. Computers & Education, 2005, 44(3):237-155.
- [3] WANG H C, LI T Y, CHANG C Y. A Web-based tutoring system with styles-matching strategy for spatial geometric transformation[J]. Interacting with Computers, 2006, 18(3):331-355.
- [4] Huseyin Koyun. Logging model for windows media services [EB/OL]. <http://www.microsoft.com/articles/LoggingModel.aspx> September.2007.
- [5] YAGER R R. An extension of the naive Bayesian classifier [J]. Information Sciences, 2006, 176(5):577-588.
- [6] HALL M. A decision tree-based attribute weighting filter for Naive Bayes[J]. Knowledge-Based Systems, 2007, 20(1): 120-126.
- [7] 令狐红英,陈梅,王翰虎,等.基于互信息可信度的贝叶斯网络入侵检测研究 [J]. 计算机工程与设计, 2009, 30(14): 3288-3290.
- [8] 吴今培.模糊诊断理论及其应用[M].北京:科学出版社, 1995.

(收稿日期:2011-05-07)

作者简介:

蒋冬梅,女,1970年生,讲师,硕士,主要研究方向:计算机网络与应用,数据挖掘。