

机器学习在适应性教学系统中的应用研究

柏宏权¹, 韩庆年²

(1 南京师范大学 教育科学学院 江苏 南京 210097 2 江苏广播电视大学 传媒艺术系, 江苏 南京 210036)

[摘要] 根据学生的学习行为主动给予适应性的教学材料是适应性教学系统研究关注的焦点问题。当前的适应性学习系统在主动适应学生的学习行为方面存在许多不足。针对适应性教学系统的这个问题, 利用机器学习方法根据学生的学习行为, 预测学生的学习风格与学习行为, 改进适应性学习系统的适应性, 提出了在适应性学习系统中使用机器学习的知识表示方法, 设计了用朴素的贝叶斯分类器动态观察和预测学生的学习行为的方案。实验结果表明, 朴素的贝叶斯分类器对于预测适应性教学系统中学生的学习行为有较高的准确率。

[关键词] 机器学习, 适应性学习系统, 朴素贝叶斯分类

[中图分类号] G40-057 **[文献标识码]** B **[文章编号]** 1672-1292(2007)04-0076-04

Application of Machine Learning in Adaptive Instructional System

Bai Hongquan¹, Han Qingnian²

(1. School of Educational Science, Nanjing Normal University, Nanjing 210097, China)

2. Department of Communication and Art, Jiangsu Radio and TV University, Nanjing 210036, China)

Abstract In the adaptive instructional system, how to give learners adaptive learning materials is a focus question. There are many deficiencies with the current adaptive instructional system's automatically adapting students' learning style and learning behavior. Aiming at this problem of the adaptive instructional system, this paper uses knowledge expression methods which fit machine learning and design a scheme by using Naïve Bayes Classifier to predict learner's behavior so that these can get learner's preferences. The result of experiment shows that Naïve Bayes Classifier has good accuracy in predicting learner's behavior in the adaptive instructional system.

Key words machine learning, adaptive instructional system, Naïve Bayes classifier

0 引言

能够根据学习者的类型与知识水平动态创建或组织信息的教学系统被称为适应性教学系统^[1]。当前, 适应性教学系统中实现适应性的主要依据是学生的学习水平与学习偏好。学生学习水平可能通过观察学生在系统中的学习行为与做测试题的表现获得。而对于学生学习偏好, 各适应性教学系统的处理方式则不尽相同。如: Carver 开发的一个适应性教学系统 CS19^[2]便是根据 Feller 的学习模型^[3]确定学习偏好的。这个系统在学生学习入口处设置不同的入口点, 每个入口点代表一种学习偏好。这样就可以根据学生的不同学习偏好给其最适合的方式来浏览适应性的学习材料。CAMELEON 系统^[4]在开始学习阶段给学生提供一个调查表, 系统通过学习者回答调查表中的一系列问题来判断他们的学习偏好。当学习者登录课程后系统根据对学习者的个性化的学习偏好评价的结果给其提供相应的学习材料。然而, 这些预设型的做法未必能准确反映出学生的实际偏好, 因为学习偏好在实际情境中会有个性化的差异, 亦会随着学习的进程而不断变化。在建构适应性教学系统中用机器学习与朴素的贝叶斯分类器动态了解学生的学习偏好, 有较高的准确率。

1 用于机器学习的内容对象描述

内容对象是显示给学生的内容信息, 是领域中的最小的教学材料单元。在本研究中, 内容对象是一些

收稿日期: 2006-08-29

作者简介: 柏宏权(1975-), 讲师, 博士, 主要从事适应性学习和适应性学习系统的教学与研究。E-mail: ba hongquan@163.com

文本片段、图片、动画、视频、Applet、ActiveX 及一些与内容呈现相关的对象。这些信息可以是适应性的, 允许不同的学生浏览不同的课程材料。

为了使内容具有适应性, 且适应机器学习的要求, 根据当前对学习偏好的相关研究结果^[5-7], 本文对内容对象进行了特殊的标注, 把每个内容对象描述成一个特征集, 如表 1 所示。

系统根据这些特征对每一个展示给学生的内容对象进行分类。前 3 个特征描述了对对象的基本特征, 由创建者给出。第 4 个特征根据对象在主题中的位置决定。使用这个属性的原因在于, 学生有不同的学习偏好, 有些学生在主题开始前喜欢某些类型的对象, 而有些则喜欢在主题的最后出现这样的对象。例如: 有些学生喜欢一开始学习摘要性的解释, 后面再学习具体的示例。有些学生则喜欢先学习具体的示例。这个属性能帮助系统决定内容对象的展示顺序。第 5 个属性与第 4 个类似, 这个属性也是决定内容对象的显示顺序。

表 1 内容对象的特征集

Table 1 Feature set of content objects

特征	值
媒体类型	{文本, 图像, 动画, 插件}
教学类型	{解释, 示例, 描述, 定义}
抽象程序	{抽象, 具体}
主题中的位置	{开始, 中间, 结尾}
概念中的位置	{开始, 中间, 结尾}
是否需要	{是, 否}

“是否需要”属性是通过观察学生是否浏览了相应的材料来决定它的值是“是”还是“否”。该属性涉及机器学习与贝叶斯分类, 可以用机器学习进行预测。

2 机器学习动态了解学习偏好并预测学习行为的实现方法

本文用朴素的贝叶斯分类器动态了解并预测学生的学习行为。选择朴素贝叶斯分类方法是因为它快速, 在现实情境中能很好工作, 且被证明对用户建模非常准确。它能在数据量很少的情况下有效地工作。设想在系统与学生进行一些简单的交互后就开始机器学习。

朴素贝叶斯分类器可以用于解决以下问题, 每一个实例 x 可通过一个联合的属性值描述且可通过一个目标方程 $F(x)$ 从有限的集合 V 呈现一些值。把属性集列表和它的相应的分类用来建造一个训练集。当一个新的实例呈现出来时, 可能根据训练集生成一个预测值。系统根据训练集给目标方程选择可能性最高的值。

朴素贝叶斯分类是用来预测学生是否呈现某一内容对象。如果学生呈现了这些对象, 那么它们的属性“是否需要”就描述为“是”, 否则描述为“否”。

为了决定要呈现的内容对象, 系统首先聚合正确难度等级上的内容对象, 把相同的特征放进特征类。如果这些对象被呈现出来了, 则这些对象被作为一个组。利用这个聚合技术, 学生可以自动得到相应难度等级的他所想要呈现出来的内容对象。

通过这个聚合, 系统检测问题中的每一个内容对象, 用朴素的贝叶斯分类器来检测这个对象是否适合呈现。朴素的贝叶斯分类器比较当前内容对象特征与实例空间中内容对象的特征, 这个实例空间包括过去学生呈现或未呈现的内容对象。

一个内容对象有如下特征:

```
<MediaType= Text
InstructionalType= Definition, Abstractness= Concrete,
PlaceInTopic= Before, PlaceInConcept= Middle>
```

存在如下公式:

公式 1

$$P(Wanted = yes) \times P(MediaType = Text | Wanted = yes) \times P(InstructionalType = Definition | Wanted = yes) \times P(Abstractness = Concrete | Wanted = yes) \times P(PlaceInTopic = Beginning | Wanted = yes) \times P(PlaceInConcept = Middle | Wanted = yes)$$

公式 2

$$P(Wanted = no) \times P(MediaType = Text | Wanted = no) \times P(InstructionalType = Definition | Wanted = no) \times P(Abstractness = Concrete | Wanted = no) \times P(PlaceInTopic = Beginning | Wanted = no) \times P(PlaceInConcept = Middle | Wanted = no)$$

每一个个体的可能性可以很容易地计算出来。一个对象呈现的可能性是这个对象在整个对象实例空间中呈现的几率。

为了计算条件概率, 本文把对象分成两个实例集: 一个“是否需要”属性是“是”, 另一个“是否需要”属性是“否”. 对于公式 1 统计“是否需要”属性是“是”时相应的属性值. 例如:

$P(\text{MediaType} = \text{graphic} | \text{Wanted} = \text{yes})$ 是图像对象在所有对象中想要呈现的次数.

$P(\text{MediaType} = \text{graphic} | \text{Wanted} = \text{no})$ 是图像对象在所有对象中不想要呈现的次数.

通过这些特征, 可以判断每一个属性对于学生的重要程度. 另外, 通过这些联合特征描述的学生实例空间 (如表 2 所示), 可以表示学生的偏好. 这样系统便可以根据学生的偏好把最适合的内容对象呈现给学生.

学生实际应用系统时, 在相应教学单元相应难度等级中呈现出来的内容对象“是否需要”属性值是“是”, 未呈现的内容对象“是否需要”属性值是“否”. 这些对象被加进朴素贝叶斯分类器的实例空间, 从而增加训练集.

表 2 实例空间的实例

Table 2 Examples of instance spaces

教学类型	媒体类型	在主题中位置	在概念中位置	抽象程度	需要
示例	图像	开始	开始	抽象	是
定义	图像	开始	结尾	抽象	是
定义	图像	结尾	结尾	抽象	是
描述	图像	开始	开始	抽象	是
示例	文本	开始	中间	抽象	是
定义	图像	结尾	中间	抽象	否
示例	文本	结尾	中间	抽象	否
示例	文本	结尾	结尾	抽象	否
解释	文本	开始	开始	具体	否

需要指出的是, 朴素贝叶斯分类器预测的“是否需要”属性值不会加入到实例空间, 只有实际值才会被加入到实例空间. 实例空间的增加将为将来的分类提供依据.

3 实验结果与分析

3.1 朴素贝叶斯分类器的准确性

本文在适应性教学系统 (I-Tutor) 中测试了分类的准确性, 检测它能在多大程度上预测学生将要浏览什么样的课程材料. 通过对 120 名学生进行测试, 发现用朴素贝叶斯分类器分类算法及对算法进行的一些阈值改变, 对每一位学生至少能获得 63% 的准确率.

用学生实例来分析朴素贝叶斯分类器的准确率, 统计所有的对象集的正确预测, 抽取了其中的 8 个学生, 如表 3 所示.

表 3 朴素的贝叶斯分类器分类的准确率

Table 3 Accuracy of using Naïve Bayes classifier

学生	准确率 %
1	84. 11
2	74. 16
3	60. 06
4	76. 97
5	84. 71
6	59. 50
7	84. 62
8	90. 53

3.2 改变阈值后的准确率

朴素的贝叶斯分类器仅仅用来决定一个对象可能被学生呈现的可能性是否大于不可能被学生呈现的可能性 (见公式 1、公式 2). 这其实就是一个 50% 阈值, 来决定是否呈现对象. 当然, 如果这个阈值变化了, 则相应分类的准确性也受到影响. 使用不同阈值的情况如表 4 所示.

通过这些对朴素的贝叶斯分类器的改进, 最好的结果见表 5

表 4 不同阈值时分类的准确率

Table 4 Accuracy of using different thresholds (单位: %)

学生	30	50	70	80	90
1	63. 08	84. 11	81. 78	79. 91	79. 44
2	67. 03	74. 16	79. 69	79. 91	80. 20
3	60. 06	60. 06	64. 09	63. 45	61. 86
4	49. 34	76. 97	76. 97	76. 97	76. 97
5	69. 41	84. 71	78. 82	78. 82	78. 82
6	60. 25	59. 50	58. 25	57. 25	58. 50
7	81. 54	84. 62	84. 62	86. 15	86. 15
8	81. 05	90. 53	91. 58	91. 58	92. 63

表 5 改进后的准确率与原始的准确率的对比

Table 5 Best improvement from original algorithm

学生	原来的准确率 %	改进后的准确率 %	t
1	84. 11	84. 11	—
2	74. 16	80. 20	6. 254
3	60. 06	64. 83	2. 733
4	76. 97	76. 97	—
5	84. 71	84. 71	—
6	59. 50	63. 00	1. 032
7	84. 62	87. 69	1. 425
8	90. 53	92. 63	1. 422

可以看出 8 位学生中, 有 2 位学生获得了显著的提高. 其它 3 位学生也有所提高. 尽管不是很明显, 但即使有一点差别也会给学生的学习带来一些改变.

4 结论

朴素贝叶斯分类对于所有学生的学习能取得较好的分类结果. 如果给每一位学生以同样的参数, 则贝叶斯分类得不到最好的结果. 通过改变域值的办法可以在一定程度上提高分类的准确率. 从结果可以看出机器学习在适应性教学系统中有着良好的应用前景.

尽管朴素贝叶斯分类对于所有学生的学习能取得较好的分类结果, 但仍然有进一步改进的空间. 还可验证其它的分类技术如: 动态决策树、支持向量机、 K 邻近等技术对于学习偏好的分类结果的准确率^[8-11].

另一个提高准确率的方法是对分类用的内容对象的特征描述作进一步改进. 因为若去除一些特征, 则会影响分类的结果. 而且, 在对内容对象表示时, 肯定有一些对于描述学生的学习很有用但被忽略了的特征.

[参考文献] (References)

- [1] Feller R M, Silverman L K. Learning and teaching styles in engineering education[J]. Engineering Education, 1988(3): 674- 681.
- [2] Carver C A, Howard R A, Lane W D. Different learning styles through course hypermedia[J]. IEEE Transactions on Education, 2004(3): 33- 38.
- [3] Feller R M, Silverman L K. Learning and teaching styles in engineering education[J]. Engineering Education, 2004(4): 674- 681.
- [4] Laroussi Mona. Providing an adaptive learning through the web case of CAMELEON[C] // Computer Aided Medium for Learning on Networks. Urbana: University of Illinois Press, 2003. 567- 571.
- [5] 赖丁·雷纳. 认知风格与学习策略[M]. 上海: 华东师范大学出版社, 2003. 47- 48.
Lartin T N. Cognize Style and learning Strategy[M]. Shanghai: East China Normal University Press, 2003. 47- 48. (in Chinese)
- [6] 谭顶良. 学习风格论[M]. 南京: 江苏教育出版社, 1995. 317.
Tan Dingliang. Learning Style[M]. Nanjing: Jiangsu Education Press, 1995. 317. (in Chinese)
- [7] Rajeev Rastogi, Kyuseok Shim. PUBLIC: A decision tree classifier that Integrates building and pruning[C] // Proceeding of the 24th VLDB Conference. New York: Cambridge University Press, 1998. 127- 131.
- [8] Kwok T Y. Automatic text categorization using support vector machine[C] // Proc Int Conf on Neural Information Processing. Beijing: Tsinghua University Press, 1998. 347- 351.
- [9] 苏伟峰, 李绍滋, 李堂秋. 一个基于概念的中文文本分类模型[J]. 计算机工程与应用, 2002, 38(6): 193- 195.
Su Weifeng, Li Shaozi, Li Tangqiu. A module of automatic chinese documents classification based on concept[J]. Computer Engineering and Applications, 2002, 38(6): 193- 195. (in Chinese)
- [10] 肖旻, 刘晓璐, 屠立忠. 基于贝叶斯分类的邮件过滤方法及模型研究[J]. 南京师范大学学报: 工程技术版, 2006, 6(2): 86- 89.
Xiao M in, Liu Xiaolu, Tu Lizhong. Research in a method and model of spam filtering based on Bayesian classifier[J]. Journal of Nanjing Normal University: Engineering and Technology Edition, 2006, 6(2): 86- 89. (in Chinese)
- [11] 王斌, 潘文峰. 基于内容的垃圾邮件过滤综述[J]. 中文信息学报, 2005, 19(5): 1- 10.
Wang Bin, Pan Wenfeng. A survey of content-based anti-spam email filtering[J]. Journal of Chinese Information Processing, 2005, 19(5): 1- 10. (in Chinese)

[责任编辑: 刘健]