

基于显性用户特征的数字电视节目推荐系统

徐江山^{1,2}, 卢增祥¹, 陶 疆², 李衍达¹

(1. 清华大学自动化系, 北京 100084; 2. 深圳市天威视讯股份有限公司, 深圳 518036)

摘 要: 为了解决数字电视系统节目“信息过载”问题, 阐述了解决该问题的数字电视节目推荐系统的背景、概念、结构和原理, 提出了基于商用播出数字电视平台和显性用户特征的数字电视节目推荐系统中节目特征、用户特征的设计思路, 提出了一种修正的基于显性用户特征的推荐算法, 设计的数字电视节目推荐系统目前已进入实际运行测试阶段。测试结果表明: 该文提出的设计与推荐机制产生的推荐结果能够符合用户的收视特性, 达到 75% 以上的正确率。

关键词: 数字电视; 推荐系统; 节目特征; 用户特征; 匹配算法

中图分类号: TN 949

文献标识码: A

文章编号: 1000-0054(2007) 01-0154-03

Digital TV program recommendation system based on explicit user profiles

XU Jiangshan^{1,2}, LU Zengxiang¹, TAO Jiang², LI Yanda¹(1. Department of Automation, Tsinghua University,
Beijing 100084, China;2. Shenzhen Topway Video Communication Co., Ltd,
Shenzhen 518036, China)

Abstract Information overload occurs with uses of digital TV systems that have a very large number of channels and programs. A Digital TV Program Recommendation System was developed to organize and simplify the data using program profiles and user profiles, with a recommendation mechanism based on the explicit user characteristics. The system is being tested at the Shenzhen commercial digital TV platforms with test results showing that the system recommendations match the users' interests 75% of the time.

Key words digital TV; recommendation system; program profile; user profile; matching mechanism

随着数字电视技术的快速发展, 有线数字电视系统在目前编码与调制制式下可以达到 500~600 套节目的传输容量。将出现电视用户在如此众多的节目中无法挑选他们感兴趣的内容的问题。为彻底解决这种电视信息“过载”问题, 电子节目指南必须

具有智能性, 它能够根据用户的兴趣、爱好和规律自动提前向用户推荐电视节目, 同时它还能自动跟踪用户兴趣的变化而对所推荐的电视节目作出调整, 这就是数字电视节目推荐系统的概念。

在电视节目推荐系统领域, TV-Advisor 系统^[1]利用用户的显性反馈自动调整用户特征, 从而向用户推荐符合其特征的电视节目; P-EPG (Personal-Electronic Program Guide) 系统^[2]和 Multi-Agent 系统^[3]在利用显性反馈信息的同时, 从用户收视记录中自动提取用户的隐形反馈信息, 从而更好更全面地反映用户兴趣特征。上述两个系统都是基于内容的推荐系统。PTV (Personal TV) 系统^[4]和 TV-Scout 系统^[5]在应用基于内容的推荐机制的同时引入合作过滤机制, 使得推荐的节目更加符合用户的心愿。

1 数字电视节目推荐系统的总体框架

基于内容过滤和合作过滤混合方式的个人电视节目推荐系统主要由以下 4 个功能模块组成: 用户特性文件模块、节目特性文件模块、合作过滤模块、推荐模块。

1.1 节目特性描述

节目的特性可以从节目分类、节目制作信息(主要演员、导演、制作商、制作年代等)、节目内容信息(对节目内容的文字描述)、节目播出信息(播出频道、播出时间、节目时间)等几个方面来描述。国家标准和 DV B(digital video broadcasting)对电视节目进行了两层分类。因此可以从 SI(service information, SI)信息中得到电视节目的分类信息。但是由于 SI 的分类信息不尽完善, 并且 SI 信息有时也无法得到, 因此, 有必要自己定义更加详尽更加深入的

收稿日期: 2005-12-28

作者简介: 徐江山(1968-), 男(汉), 湖北, 博士研究生。

通讯联系人: 卢增祥, 副教授, E-mail: luzx@tsinghua.edu.cn

分类树。另外,由于同一节目可能会属于不同的类别,还必需定义每一个节目属于不同类别的权重。分类信息只能框架性地定义节目属性,如果需要进一步描述节目特性,文字描述是有益的。对节目特性的描述涉及不同的描述方式,例如有数字型描述(如节目时间等)、文字型描述等,因此在与用户特性文件的匹配时必需考虑描述方式的对应,并且不同的描述方式需要采用不同的匹配准则(例如文字型的采用 Boolean 准则)。

1.2 用户特性描述

用户特性文件描述用户对电视节目的喜好、不喜好和要求。由于需要反映用户对电视节目不确定的需求或喜好,所以对用户特性的描述必需用到模糊集的概念。另外,由于电视节目所包含的领域广泛,用户对电视节目的喜好会发生变化,为动态地捕捉用户的这种变化,必需引入反馈的概念,使用户特性文件自适应地随用户兴趣的变化而变化。

初始用户特性的获取有 2 种方式: 第 1 种方式是在用户注册成为推荐系统的用户时要求用户向系统提供系统设计的调查问卷; 第 2 种方式是根据社会调查信息给予用户初始特性。必要时需要将两种方式结合起来。在第 1 种方式中,用户需要提供的信息分为两类: 第 1 类是用户的性别、职业、年龄……、收看电视的时间、喜好节目的类别等硬性信息; 第 2 类是用户曾经看过的最喜欢的电视节目和最不喜欢的电视节目(数量越多越好)等软性信息。

用户特性文件的自动更新依赖于用户信息的反馈和反馈信息的提取。用户的信息反馈分为显性反馈和隐性反馈两种。显性反馈是系统向用户提供交互操作界面,用户可对系统提供的推荐结果作出评价,并且用户随时可以修改其特性文件的内容。显性反馈有利于反映用户突然的兴趣变化,对于非交互式电视网络(没有使用 STB),显性反馈是唯一的反馈方式。隐性反馈依靠 STB 自动记录用户的历史收视清单,并将这些信息自动反馈给系统以提取有用的信息,隐性反馈更有利于反映用户渐进式的兴趣变化。

1.3 推荐机制

个人电视节目推荐系统可以采用基于内容相似度匹配的推荐机制,也可以采用基于合作过滤的推荐机制。基于内容的推荐机制通过计算用户特征向量与节目特征向量之间的相似度,然后将相似度高的节目推荐给用户。在计算相似度的过程时,还必须

考虑分量特性在相似度计算中的权重,基于内容的推荐方法除了利用上述相似度匹配方法之外,也可以采用 Bayes 分类算法。由于电视节目内容描述信息有限,因此基于内容的推荐机制不能充分匹配用户和节目的特性。合作过滤机制通过寻找与特定用户有相似爱好的 k 个近邻,将系统推荐给这 k 个近邻的电视节目中具有较高频率的节目也推荐给该用户。合作过滤的关键在于近邻的选取,但是近邻的选取要求该用户具有较长时间的收视记录,因此,对于新注册用户,系统还必须依赖基于内容的推荐机制,同时对于用户兴趣的突然变化,合作过滤推荐机制也无法作出及时的反应。所以,完善的个人电视节目推荐系统必须将基于内容的推荐机制和基于合作过滤的推荐机制有机地结合起来。

2 系统设计

作为数字电视节目推荐系统研究的第 1 步,构建一个结构完整、基于实际播出平台的系统是必要的,本文作者依托深圳市有线数字电视平台真实的频道、用户和节目环境,基于显性用户特征,搭建了国内第 1 个数字电视节目推荐系统。

2.1 用户显性特性文件的定义

用户在申请注册成为系统用户的过程中,系统将要求用户提供两方面的用户信息。第 1 类信息是普通用户信息,例如姓名、年龄、性别、职业、收入水平、受教育程度、订购的频道名称、希望的节目推荐方式(邮件、手机短信、网站主页)等; 第 2 类信息是用户的收视喜好信息,例如对各频道的喜好程度、对各收视时间段的喜好程度、对节目类别(包括二级类别)的喜好程度、从系统所列演员和导演库中挑选所喜爱的对象并注明喜好程度等。第 2 类特征是一种定量的用户特征,将用于推荐算法的计算。

2.2 节目特性文件的定义

对于一个特定的节目,其特性文件由下列要素组成: 节目名称、节目播出的时间以及持续时间、节目播出的频道、节目第 1 和第 2 级类别、演员或主持人信息、导演或制作人信息、节目内容简介、节目剧照或片花等。在规划节目特性文件过程中,我们又将其分为动态信息和静态信息两种,动态信息主要指各频道节目的名称及播出时间,静态信息主要指节目对应的其他信息,并且采用 XML 格式对节目特性进行定义。以下是用 XML 格式对中央电视台 2005 年 11 月 21 日晚上 19:55 至 20:47 时段节目特性文件的定义示例:

```
< Program channel= "13" start= "20051121195500" stop=
"20051121204700">
< title lang= "zh-chs">剧场: 家风 ( 2)< /title>
< credits>
< actor>王庆祥< /actor>
< actor>王海燕< /actor>
< /credits>
< desc lang= "zh-chs">锦萍偷偷跑到 幼儿园接出了贝贝!
杨正民终于在游乐场见到了自己的亲孙女……< /desc>
< Content Type>
< BaseType>电视剧< /BaseType>
< Extended Type>都市生活情感< /Extended Type>
< /Content Type>
< /Program>
```

在 DVB分类与国家分类标准之上,我们结合中国的实际情况,对节目分类进行了调整,定义了电影、电视剧、新闻、财经、文化娱乐、体育健身、戏曲、少儿、科教、动画、纪录片、旅游生活、访谈、军事、法制、专题 16个 1级类别,和 124个 2级类别。例如电视剧的 2级类别有港台、日韩、外国、政治题材、历史题材、都市生活情感题材、公检法涉案题材、农村题材、少儿题材、部队生活题材、宫廷戏说题材、武侠神话题材、室内情景喜剧、电视艺术片和动画片 15个。

2.3 推荐机制

文[6]阐述了基于用户显性特征的推荐方法。显性推荐指数为

$$E = w_t r_t + w_c r_c + (1/K) \sum w_i r_i.$$

w_t 、 w_c 、 w_i 分别表示观看电视喜好的时间、频道和属性(属性有 K 类)的权重,分别取 0.1、0.2和 0.7; r_t 、 r_c 、 r_i 分别表示对应的值。

本文基于上述节目特性的实际情况,对上述算法作了调整,具体表示如下:

$$E = w_t r_t + w_c r_c + w_g \sum \frac{w_i r_i}{w_i}.$$

式中的 w_g 表示属性项的整体权重,时间、频道与属性的权重仍然按照文[6]的分布,各子类属性的权重将视子类权重的重要程度确定。

3 测试结果

遵循上述设计方法,完成了系统总体设计、数据库结构设计、节目特征文件与用户特征文件的采集和格式定义、算法设计以及界面设计。系统投入试运行,系统网站为 211.148.195.30/mytv/。

上述系统能够实现全部频道一周内播出节目的显示与查询;能够根据用户注册提供的收视特性提供符合用户收视需求的当前时段与全天时段的节目推荐,全天时段除了在推荐页面提供 TOP20即 20个最符合用户喜好的节目外,还通过 More按钮提供 TOP50;另外,系统能够提供用户对推荐结果的反馈机制,系统设计了非常喜欢、喜欢、一般喜欢、不喜欢和非常不喜欢 5级反馈机制。基于上述系统,我们随机选取了 10位实验用户,并采用推准率的评测机制,对系统的性能进行了评测,测试结果见表 1。

表 1 系统测试结果

反馈类别	反馈个数	比例
非常不喜欢	9	4.25
不喜欢	38	17.92
一般喜欢	28	13.21
喜欢	66	31.13
非常喜欢	71	33.49
合计	212	100.00

4 结 论

测试表明,系统推荐结果能够符合用户的收视特性,推准率(反馈为一般喜欢和非常喜欢时,系统认为推荐准确)达到 75%以上。

参考文献 (References)

[1] Das D, Ter Horst H. Recommender systems for TV [C]// Recommender Systems, Papers from the 1998 Workshop. Madison, CA: AAAI Press, 1998: 35-36.

[2] Ehrmantraut M, Hürder T, Wittig H, et al. The personal electronic program guide—towards the pre-selection of individual TV programs [C]// Proceedings of the 5th International Conference on Information and Knowledge Management (CIKM'96). Maryland USA: ACM Press, 1996: 243-250.

[3] Kurapati K, Gutta S, Schaffer D, et al. A Multi-Agent TV Recommender [EB/OL]. 2001. <http://www.di.unito.it/~liliana/UM01/kurapati.pdf>.

[4] Cotter P, Smyth B. PTV: Intelligent personalised TV Guides [C]//Seventeenth National Conference on Artificial Intelligence. Austin, TX, USA: AAAI Press/The MIT Press, 2000: 957-964.

[5] Baudisch P. Dynamic Information Filtering [D]. Sankt Augustin: Forschungszentrum Informationstechnik GmbH, 2001.

[6] Zimmerman J, Kurapati K, Schaffer L. TV personalization system: Design of a TV show recommender engine and interface [C]//Personalized Digital Television. Netherlands: Kluwer Academic Publishers, 2004: 27-51.