

文章编号: 1001-9081(2013)S2-0119-06

国外电影推荐系统网站研究与评述

孙海峰, 甘明鑫*, 刘鑫, 吴越

(北京科技大学 东凌经济管理学院, 北京 100083)

(* 通信作者电子邮箱 ganmx@ustb.edu.cn)

摘要: 电影推荐网站的研究对于解决电影数据复杂性以及用户需求多样性具有重要的意义。针对电影推荐系统中推荐准确率低、数据稀疏和冷启动等问题, 详细阐述了国外六大主流电影推荐网站系统的内部结构、信息量和推荐方式; 重点阐述了各网站的推荐异同以及可优化方案, 进而通过新用户对电影进行评分的冷启动实验, 详细分析冷启动解决方法与效果; 最后指出当前电影推荐网站的现状及问题, 对推荐系统相关理论及其发展趋势进行了分析和展望, 并为实际应用中解决冷启动与数据稀疏等问题提供参考依据。

关键词: 电影推荐系统; 协同过滤; 冷启动; 数据稀疏; 标签

中图分类号: TP393 **文献标志码:** A

Review on dominating websites for movie recommender systems

SUN Haifeng, GAN Mingxin*, LIU Xin, WU Yue

(Dongling School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China)

Abstract: The research and practice of movie recommender system is important in solving the complexity of movie data and diversity of user requirements. Focused on the problems of low accuracy, data sparsity and cold start in six mainstream foreign movie-recommend websites, this paper summarized the systematic internal structure, the detailed recommended way, the movie information, etc. This paper also emphasized on the similarities and differences of the recommendation ways among the websites and the optimized proposal; and furthermore, according to the grading of movies by new users, the cold start experiment analyzed the way and effect of cold start solutions. Finally, the paper clearly indicated the situation of current movie recommendation websites. The opening issues and challenges for the movie recommendation system concerning both the theory and trend were proposed, which provide a reference for practical application in the solution of cold start and data sparsity.

Key words: movie recommendation system; collaborative filtering; cold-start; data sparsity; tag

0 引言

由于 Internet 具有信息传输的开放性、信息交流的双向性, 以及信息覆盖的广泛性等特点, 网络中的信息量以指数规律迅速扩展和增加, 网络上的“信息过载”和“信息迷航”问题日益严重。解决这些问题的关键在于将 Internet 从被动接受浏览者的请求转化为主动感知浏览者的信息需求, 实现网络系统对浏览者的主动信息服务。为了给人们提供满意的信息和服务, 推荐技术应运而生, 成为目前众多学者和网络用户关心的核心技术。推荐系统通过预测用户对信息资源的喜好程度来进行信息过滤, 根据用户具体需求通过协同过滤等技术进行个性化推荐。根据这一基本原理, 推荐系统产生了各种不同算法。在众多推荐算法中, 协同过滤推荐是迄今为止最为成功也是应用最广泛的个性化推荐技术。协同过滤的概念被广泛应用于各个领域, 如电子商务、社交网络、图书推荐、电影推荐等。国外比较著名的推荐系统有 Amazon.com, GroupLens 和 Ringo 等; 国内的推荐系统近几年也处于一个高速发展的时期, 如淘宝、京东商城等。

个性化推荐技术发展的同时, 也避免不了本身固有的问题, 如数据稀疏和冷启动问题, 以及推荐精确性问题。数据稀疏问题^[1]是指系统中一般用户购买的物品总量或评分的物品

总量占网站物品量的 1% 左右, 造成评价矩阵 (用户—物品矩阵) 非常稀疏, 从而难以找到最近相邻用户集, 使相似度计算耗费增加。冷启动问题从一定角度可以看成是稀疏问题的极端情况, 指新物品首次出现后未得到任何用户评价或者新用户进入系统后未对任何物品进行评价而造成难以推荐和推荐精确度差的问题^[2]。因此在保证精确性的前提下, 降低数据稀疏和冷启动问题的影响成为当前个性化推荐的热门方向。

近几年来, 标签在各大推荐网站被广泛使用, 使其成为一种有效的推荐工具。标签是附加到数字对象上用来描述该对象的关键词, 但并非正式分类系统的一部分。它是被自由选择的关键词, 却是一种用来组织、搜索、和探索对象资源的简单而强大的工具。它有很多优点: 首先, 它更直接地反映出用户对一部电影的感受。区别于用户评论, 标签没有严格的组织结构, 却包含丰富而明确的主题信息^[3], 从而避免了用户评论的繁琐; 同样, 区别于电影类型信息, 标签能够更灵活地存储用户感受; 另外, 标签行为形成了一个基于用户、对象、标签的三维关系, 除传统的 user-item 关系, 增加了 item-tag^[4] 以及 user-tag^[4] 的对应关系。由于这些优势, 标签越来越受欢迎, 并被广泛应用于个性化推荐。

本文将推荐技术着眼于电影推荐, 主要介绍和分析 Jinni^[5]、IMDB^[6]、Criticster^[7]、Movielens^[8]、Nanocrowd^[9]、

收稿日期: 2013-05-29; 修回日期: 2013-08-01。

基金项目: 国家自然科学基金资助项目 (71101010); 中央高校基本科研业务费资助项目 (FRF-BR-11-019A)。

作者简介: 孙海峰 (1991-), 男, 山东青岛人, 主要研究方向: 基于本体与异质复杂网络的推荐系统; 甘明鑫 (1978-), 女, 北京人, 副教授, 博士, 主要研究方向: 信息推荐系统、复杂网络; 刘鑫 (1992-), 女 (回族), 河北沧州人, 主要研究方向: 项目管理; 吴越 (1992-), 女, 江苏无锡人, 主要研究方向: 推荐系统标签问题。

RottenTomatoes^[10] 六大主流电影推荐网站,并对六大网站的电影基本信息进行汇总,详细介绍并比较分析各网站的电影检索与推荐机制以及网站特色。进而,本文针对协同过滤推荐中的冷启动问题,对两个推荐机制较强的网站进行简单实验并对结果进行分析。

1 推荐网站介绍

电影信息检索是电影网站的基本功能,即按一定的方式组织信息,并根据用户的需要找出有关信息的过程和技术。狭义的信息检索就是信息检索过程的后半部分,即从信息集合中找出所需要的信息的过程,也就是常说的信息查寻,使用户能够通过相关关键字进行简单的搜索,以获取自己明确寻找的电影。电影的推荐系统则根据用户的兴趣特点和行为,向用户推荐其感兴趣的信息和电影。表 1 列出了各网站的电影基本信息。

表 1 网站基本信息表

电影网站	电影信息
Jinni	主演,导演,基于小说的作者,语言,国家,时长,美国电影协会排名,IMDB 得分,主要介绍,摘要以及用户评论等
IMDB	影片演员、导演,剧情,时长,类型,上映日期,影评,用户评分,电影排名以及影片相关的琐事花絮,片中出现的漏洞,影片音轨,屏幕的高宽比,影片的不同版本等,约 214 万部
Criticker	内容简介,电影的类型,国家,导演编剧,以及用户评分等,约 6 万部
Movielens	主演、导演、类型、语言,用户评分,平均得分,影评人数以及该部电影的标签
Nanocrowd	电影摘要,主演,演员,基于小说作者,制片商,评论(外部用户评论以及 Amazon 用户评论),美国电影协会排名,电影标签等
Rotten Tomatoes	电影简介,上映日期,排名(tomato 排名和 Audience 排名),时长,类型,导演,票房,作者,主演,评论

1.1 Jinni

Jinni 可以依据电影的情节、时长、发生地点、所获奖项、关键字等一系列参数进行搜索,功能比较完善。Jinni 所采用的检索工具叫作 Movie Genome。Movie Genome 知名度非常高,Google TV 也采用这个工具实现个性化检索。Jinni 的 Movie Genome 包含 2200 种不同的参数,包括类型、背景、氛围等。它根据用户的评论以及其他信息找到确切的参数来描述一部电影。Jinni 利用 Movie Genome 包含的大量参数将电影进行分类并且探索出用户偏好以及用户之间的关联程度。

首先用户要注册一个账户,生成个人文件夹,用来收藏用户行为及兴趣信息。用户要对至少 10 部电影进行打分,Jinni 才能做出推荐。

1.1.1 用户及电影分组

Jinni 把用户分为 12 类,分别是社会政治型、头脑聪慧型、英雄主义者、个人主义者、战略家、现实主义者、理想主义者、极端主义者、富有想象力的艺术家、戏剧爱好者、情景剧爱好者和悬疑爱好者。通过对用户的明确分类,Jinni 可以针对每一种分类从用户兴趣及行为中获取用户对该类型的偏向,从而进一步明确用户的偏好类型。而用户对电影的评级分为 10 级,由低到高分别为可怕、差、乏味、令人失望、一般、还可以、好、非常好、令人惊叹、绝对不能错过。避开传统的分数,而使用情绪评分更能够明确表现出用户对某部电影的兴程度,消除传统评分中保守评分用户兴程度模糊的问题。例如:保守评分用户的 6 分便已经代表对该电影很满意,相当于一

般用户的 8 分甚至以上,便产生偏好定位的错误。但是利用情绪评语,每一个等级代表明确的喜好程度,从而解决了前面的问题。

1.1.2 偏好定位

推荐系统根据用户的评分及行为信息对用户的数据进行分析,得到该用户的偏好,从而进行推荐。根据某用户对一部电影的评级,Jinni 可以了解到该用户是否喜欢该部电影;接着对用户电影偏好进行草图拟定。Jinni 将观看电影的人群分为 12 个大类,基于每一大类中的用户打分情况,Jinni 用六个不同的词条来刻画用户偏好。如:strong leaning 表示用户强烈倾向该类型,而 not at all 表示用户完全不是该类型。对于用户偏好的拟定,用户需要对同一大类下的至少 20 部电影进行打分,否则难以给出结果。因为在一个分类中对用户判定兴趣度十分重要,错误的评判将直接将用户移出该类型,造成后期的推荐出现极大偏差。已有的电影评分越多,越能够更完整测试出用户对本类型电影的兴趣,从而提高系统对用户兴趣的判断,避免发生偏好错误。

最后,根据与其他用户间的联系比较,用户可以了解自己与其他用户的关联程度,并且选择是否跟随。如果用户选择跟随,被跟随者最近观赏过的电影会对用户推荐产生影响。Jinni 在用户文件夹中会显示联系紧密的其他用户最近观赏过的电影以及大多数人对该电影的评分。当用户进入 Jinni 推荐页面,根据用户偏好,产生用户的推荐列表;另外根据相似用户群的最近活动信息,将评分较高的电影选出以产生另一个推荐列表。

1.1.3 显示推荐结果

完成用户兴趣信息搜集并对信息进行处理后,推荐的最后一步就是根据分析的结果为用户实现推荐。而推荐列表中每一部电影与该用户的兴趣度联系也是有差异的。在 Jinni 中搜索一部电影时,Jinni 会根据该电影相关信息,提供相关联的其他电影。推荐的电影以图片格式显示,图片大小代表对电影的排序,默认情况下是以相关性进行排序的。如:在搜索栏中搜索“The Godfather”,搜索的结果中 The Godfather 图片最大,The Godfather Part II 和 The Godfather Part III 次之,The Youngest Godfather 等图片便很小。因此,电影图片越大,代表搜索信息的相关性越强。当然,用户可以对这些电影重新排序,其中包括影片上映时间、片长以及得分的排序。针对推荐列表,用户可以再次进行信息调整,以实现二次筛选,信息调整包括电影知名度、电影严肃度、影片真实度等。

1.2 IMDB

IMDB 使用了多重检索,即同时使用多种关键字对需求信息进行筛选。IMDB 可以分别对标题、摘要、类型、关键字、电影公司、情节等进行检索,同时还具有高级搜索。高级搜索包括高级标题搜索、高级名称搜索,以及协同重叠搜索。高级标题搜索可以将电影类型、时代、投票数、评分等多重因素考虑在内进行搜索推荐。高级名称搜索可以根据影片中演员的性别、出生日期、名字以及身高等因素进行搜索。协同重叠搜索则可实现两项标题或两项名称同时检索。在标题协同搜索中同时输入“Apocalypse Now(现代启示录)”和“The Godfather(教父)”,会搜索到两部电影中同时出现的人,演员 Robert Duvall 和导演 Francis Ford Coppola 等。在名称协同搜索中同时输入导演 Frank Darabont 和演员 Tim Robbins,就会搜索到两人合作的影片,如《肖申克的救赎》等。

1.3 Criticker

Criticker 的电影评分方式很有特色,它运用兴趣兼容指

数(Taste Compatibility Index, TCI)进行用户偏好定位。当用户建立自己的账户后,可以为网站内的电影进行评分。针对用户的电影评分,Criticker 会再对评分电影自动设置等级分,范围从 1 至 10。系统的等级分是将该用户所有的电影评分排序,然后由高到低分别给予等级分数。因此,即使两个用户对同一部电影给予相同的分数,等级分并不一定相等。例如:A 用户对 1~5 五部电影的评分分别为 65、75、85、90、95,而 B 用户对相同五部电影的评分依次为 65、55、45、35、25。系统给予 A 用户的电影等级分依次为 1、2、3、4、5,而给予 B 的为 5、4、3、2、1。因此 A、B 用户对第一部电影的评分虽然相等,但是系统的等级分数却相差很大。通过这种方式,每一个用户都会产生自己所评电影的等级库,任意两个用户对于相同电影的等级分数的不同会产生差值,将所有差值平均便可得到两用户的 TCI;TCI 指数小于等于 3 时说明两用户的兴趣度较相似,因此在推荐电影时,会将对方的喜好考虑在内;TCI 大于 3,则两用户之间可认为无共同兴趣度,因此双方的选择及偏好不会彼此影响。

Criticker 使用 TCI 进行推荐的方式如下:

用户进行信息注册形成自己的账户,才可使用 Criticker 的推荐系统。当然这需要用户去选择性地对一些电影进行评分,便可形成自己的 TCI 库。

拥有自己的 TCI 库后,用户可以进入 Criticker 提供的“Best TCIS”中查看当前与该用户兴趣度相似的其他用户列表(Criticker 中称为 my top user)。当用户选择系统推荐电影,系统便会基于用户的评分电影与 TCI 对其他电影评分。这里的评分通过 PSI(The Probable Score Indicator)产生。PSI 通过将 Criticker 中某用户的 1000 个 top users 按照 TCI 排序,再将评分过某部电影的前 10 位用户的等级分数平均,成为该用户对该电影的等级分数,再根据该用户的评分电影,最终将平均等级分转变为一个具体的分数作为该用户对该电影的预测评分,然后将得分高的进行推荐。这里,并不是任何一部电影系统都会自动产生 PSI,必须是在某用户的前 1000 名 top uses 中至少 3 位对该电影评分才会形成该部电影的 PSI。

进入“PSIs”选项,可以得到最近产生的 PSI 排名前 50 的电影列表,当然,PSI 只是系统替用户做的预测,用户依然可以预测过的电影进行实际评分。

可见 Criticker 使用的推荐方法较新颖。Criticker 在预测评分时并非直接引用用户评分,而是简单巧妙地追加等级分,削弱不同用户因评分区差异过大而造成的偏差。虽然生成 TSI 时 Criticker 仅使用等级分数的简单平均,但 Criticker 为减小误差,将用户的评分电影考虑在内后将平均等级分数转化为具体分数。然而,这其中依然存在问题,如:基于用户的电影评分将等级分转化为具体分数的方法是否准确,考虑什么因素,如何提高准确性等,这些需要进一步研究。

1.4 MovieLens

MovieLens 提供三种方式搜寻电影。第一种方法是标题搜索,这种搜索得到的电影不会考虑用户是否对其评分或是否对其有预测。第二种方法是类型搜索,通过电影的类型进行检索,类型搜索不会返回用户已经评分的电影或没有预测过的电影。第三种方法是高级搜索,它帮助用户改善搜索条件,而且并非简单的标题搜索加上类型搜索。它的搜索拥有不同的方式,比如标题短语搜索、日期限制搜索、语言范围搜索等。通过高级搜索对前两种基本搜索方式的完善,用户可以基本实现模糊信息下的电影搜索。

电影推荐系统中,MovieLens 根据用户对一部分电影的评

分,预测该用户对其他电影的评分。当新用户进入 MovieLens,需要对至少 15 部电影评分,评分范围为 1~5 分,间隔 0.5 分。系统对用户的评分情况进行分析,并将结果反馈给用户,以便用户更了解自己的评分情况。在 MovieLens 的分析结果中,它会告知用户评分的电影中其他用户评分次数最少的前 5 部电影,用户未评分而其他用户评分次数最多的前 5 部电影,用户评分低于平均分的前 5 部电影以及用户评分高于平均分的前 5 部电影。另外,系统也会统计每一个得分区间用户评分的电影数目和评分数目级别上的电影数目。MovieLens 也会根据电影类型统计每一种类型的电影数目以及每种类型的平均得分。通过分析,用户可根据自己的偏好做适当调整,使推荐更准确。当用户查看某部电影,MovieLens 将根据用户以往的评分预测其对该电影的评分,然后根据电影相似性以及用户相似性建立联系,实现推荐工作。对于推荐的电影,MovieLens 同样可根据一些标签对推荐列表进行调节,使推荐更加准确。MovieLens 的优点是用户只需要对一部分电影评分,系统便可以自动为用户预测其他电影得分,一定程度上解决了初始用户数据稀疏情况下的推荐偏差问题。

1.5 Rotten Tomatoes

Rotten Tomatoes 会追踪所有的评论内容(一些主要的大型电影能达到约 250 篇)以及正面评价的比例。若正面的评论超过 60% 以上,该部作品将会被认为是“新鲜”(fresh);相反的,若一部作品的正面评价低于 60%,则该作品会被标示为“腐烂”(rotten)。电影评价分为两种,T-Meter Critics 和 Top Critics,前者是大众和专业影评人员的综合评分,后者只是专业影评人员的评分。

这样可以很容易地看出 Rotten Tomatoes 的推荐特色。第一是简单的答案,只能选择“好”或“烂”,没有分数的概念,最终的结果根据选择“好”的百分比得出。这个算法的优点是在标准比较绝对的情况下,每一张选票的好恶都很明显,没有含糊。第二就是三因素评分方法:一部电影有 3 个新鲜度,分别由普通用户、专业影评人和著名影评人评出。其中,由大量专业影评人评出的那个分数最具参考价值,也是 Rotten Tomatoes 最值得看的内容。

1.6 推荐方式比较

在这五个网站中,单从推荐方式来看,Jinni, Criticker, MovieLens 的推荐都比较有特色。三者都是要求用户在自己账户中对部分电影评分,Jinni 根据用户的评分将其分组,由此进行相关性推荐;Criticker 根据用户对电影的评分,与其他用户形成 TCI,由此进行用户联系及相关推荐,并在推荐过程中使用 PSI 方式对用户未评分电影进行预测性评分。而对于 MovieLens,用户同样需要对至少 15 部电影评分,根据评分信息,系统将自动对其他电影评分,然后将所有评分与其他用户进行相似性关联,通过算法求得相似性进行推荐。

虽然三者的算法都不相同,但都通过用户与电影的关系,再根据电影与电影的关系,最终确定用户与用户的关系,从而形成用户与其他电影的关系列表,即推荐列表。而且 Criticker 和 MovieLens 都会通过用户对少量电影的评分去预测用户对其他电影的分数,排序后实现推荐。上文已经详细描述 Criticker 的整个预测评分过程。MovieLens 并没有将其预测性评分方法公开,因此本文很难得到内部运算方式。但是单从结果看,本文对 10 部电影评分后,Criticker 最后推荐的电影数目很少,因为它在信息不足时不会对某些电影评分。而 MovieLens 会推荐一个较大的电影列表,当然此列表会根据预测评分排序,且前几部电影得分都很高。根据 Criticker 和

电影地图与标签地图不是完全独立的,在 Nanocrowd 中,二者在功能上联系到了一起:即当用户在电影地图上点击电影名称旁边的红色标志时,不仅可以浏览该电影的详细信息,还会列出此电影的相关标签,单击一个标签,便会直接进入标签地图;同样,在标签地图中,当点击标签框中的红色标志,也会展示关于此标签的相关电影。通过这种方式,定位沙盒中的两项功能就可以实现彼此的转换。这是 Nanocrowd 中比较富有特色的标签与推荐的结合。

3 数据稀疏与冷启动问题

目前,由于在实际的协同过滤推荐系统中,获取足够的用户信息以及资源信息很难在短期内完成,因此,现有大多数协同过滤推荐系统存在着两个主要问题:数据稀疏和冷启动问题。

3.1 网站分析

在协同过滤推荐中,系统通过用户对电影的评分形成推荐。对于一部新电影,当还没有任何用户对其评分及评论时,就无法实现对该部电影的推荐,即资源冷启动问题^[11];新用户进入系统中,由于没有任何浏览信息以及用户相关信息,也无法实现个性化推荐,即用户冷启动问题^[12]。这六大推荐网站并没有很好地解决两种冷启动问题,大多冷启动的推荐仅局限于常见的最受欢迎电影推荐或者随机推荐。Jinni 与 Criticker,在处理冷启动上,倾向于随机推荐;而 Nanocrowd,由于其本身就是基于内容和标签的推荐机制,因此没有冷启动问题;MovieLens 直接避开用户冷启动问题,用户注册后必须进行电影评分才有推荐功能,这就能在第一时间获取新用户的信息,从而将冷启动问题转化为数据稀疏情况下的推荐。

3.2 冷启动实验

数据稀疏是指新用户或者新资源已经有了相关信息,但是信息极少,不足以实现精确的推荐,这也是冷启动的一种情况。这里使用 MovieLens 和 Criticker 来进行比较。针对数据稀疏的推荐问题,本文做了一次推荐实验。在 MovieLens 与 Criticker 中,作为新用户进入后对同样的 15 部电影进行相同的评分,然后看两个网站的推荐结果。

在电影评分时,假设某用户喜欢“romance”“fantasy”“comedy”这些类型的电影,而不喜欢“crime”“thriller”“horror”这些类型的电影。如图 3~4 所示。

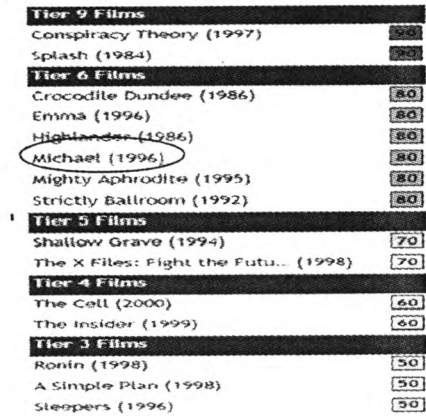


图 3 Criticker 评分

对评分数据进行的分析结果如图 5~6 所示。图 5 显示出每一种电影类型被评分的次数,图 6 则显示出在评分中每一种类型的平均得分。

根据评分电影,本文针对这两个网站分别进行对其他电影的预测评分,并将得分较高的电影推荐给用户。图 7 为 MovieLens 推荐结果,图 8 为 Criticker 推荐结果。



图 4 MovieLens 评分

从推荐结果看到两个网站有较大差异。Criticker 将其他相邻用户对某部电影的等级分数取平均,再根据用户的电影评分,从而实现推荐。一部分得分来自其他用户,另一部分来自用户本身。由此看来,Criticker 完全根据得分的高低来判断用户偏好。

仅从推荐结果来看,MovieLens 给用户推荐更多的是该用户所不喜欢的“crime”,“thriller”类型的电影,进一步分析发现,虽然本文对“crime”“thriller”类型的电影打了较低的分,但是在有评分的电影中,这些类型电影数目较多。MovieLens 认为用户评过分的电影表示用户已看过,也即用户对这些电影更感兴趣,而不论看后是否满意。那么 MovieLens 更侧重于用户的关注度,在推荐时对于用户关注过并评低分的电影仍然会认为用户对其感兴趣。Criticker 和 MovieLens 的两种推荐结果各有特色,Criticker 更看重得分,MovieLens 更看重用户的浏览次数与用户行为。

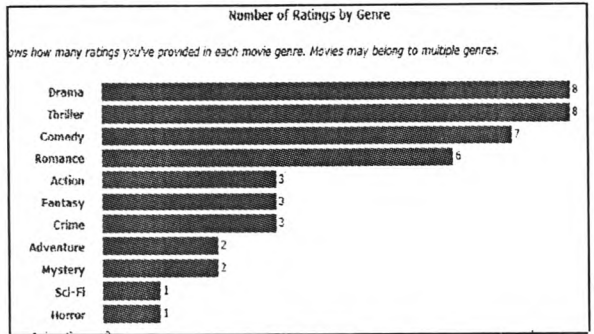


图 5 电影类型评分次数

4 结语

本文通过对六大主流电影推荐网站的具体分析,对电影信息汇总,详细介绍了各网站电影检索与推荐方法,并分析各电影

推荐网站的特色之处。最后针对协同过滤中的冷启动和数据稀疏问题,通过实验具体分析比较两个网站的电影推荐异同。

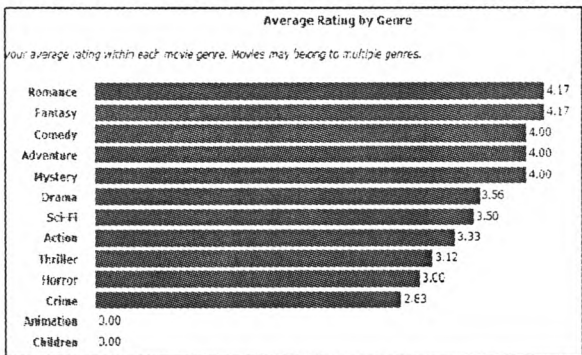


图6 电影类型平均得分



图7 MovieLens 推荐结果

目前协同过滤的推荐系统并没有完全成熟,单从电影推荐网站来看,其问题如下:许多电影推荐网站仍然停留在信息检索层面,并没有做到真正的个性化推荐,也没有解决用户特定需求的个性化问题;对于个性化推荐,大多电影推荐网站也没有应用有效的方法来解决冷启动和数据稀疏等主要问题,或其方法准确性不高;在电影信息冗余情况下,实现协同过滤时提取有效信息的准确性不高。要较好地解决冷启动问题,应考虑如下几个方面:首先,要解决新电影的冷启动问题,即如何实现对新增电影的精确推荐。推荐系统要在电影原始信息之外,更多更快地增加用户添加的信息,而这就需要相应地增加用户对新电影的评价机会。其次,对于新用户问题,可以在用户注册过程中增加用户兴趣导向等基本属性,使系统能在新用户注册时立即获取兴趣信息,从而在用户冷启动时能够根据兴趣匹配实现简单推荐。而实现这种推荐方式,需要构造出“用户—资源”模型来连接用户与兴趣资源。对于新用户,可以根据一些重要的属性特征,将其分配到相对应的用户模型;同样对于新电影,利用贝叶斯分类方法,将其分配

到对应的电影类别模型。另外,标签具有的独特优点,使其成为推荐系统的一个发展方向,有效利用标签传递的简单有力的信息,既可以计算用户相似度,也可计算电影相似度,从而建立用户—标签—电影三者的数据模型,实现最终的推荐。

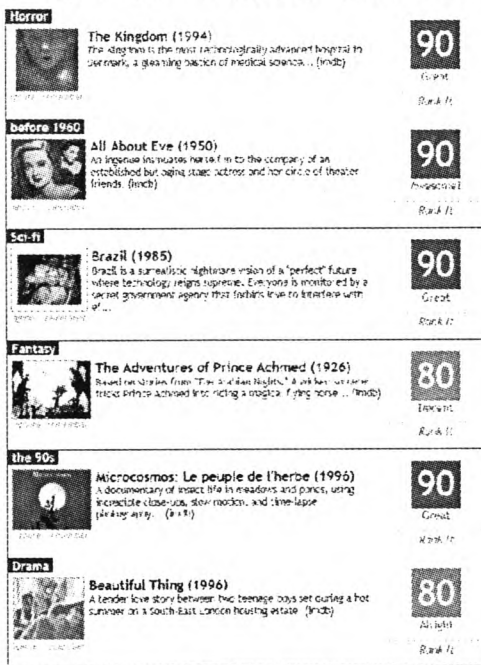


图8 Criticker 推荐结果

要提高未来的电影推荐系统的准确性,并解决存在的主要问题,就必须加强有效协同过滤方法的使用,并更加重视个性化推荐中冷启动、数据稀疏等问题,而在计算用户相似度及电影相似度方面,需要在信息选择上更加慎重。标签以其简短明确以及用户与电影的关联性强等优点,可能将实现电影推荐的重大突破。

参考文献:

- [1] 吴颜,沈洁,顾天竺,等. 协同过滤推荐系统中数据稀疏问题的解决[J]. 计算机应用研究, 2007, 24(6): 94-97.
- [2] EMBARAK O H. A method for solving the cold start problem in recommendation systems[C]// Innovations in Information Technology. Abu Dhabi: IEEE, 2011: 238-243.
- [3] ZHOU X J, XU Y, LI Y F. The-state-of-the-art in personalized recommender systems for social networking [J]. Artificial Intelligence Review, 2012, 37(2): 119-132.
- [4] ZHANG Z K, LIU C, ZHANG Y C, et al. Solving the cold start problem in recommender systems with social tags [J]. Europhysics Letters, 2010, 92(2): 28002.
- [5] Jinni [EB/OL]. [2013-05-20]. <http://www.jinni.com/>.
- [6] IMDB [EB/OL]. [2013-05-20]. <http://www.imdb.com/>.
- [7] Criticker [EB/OL]. [2013-05-20]. <http://www.criticker.com/>.
- [8] MovieLens [EB/OL]. [2013-05-20]. <http://movielens.umn.edu/login>.
- [9] Nanocrowd [EB/OL]. [2013-05-20]. <http://www.nanocrowd.com/>.
- [10] Rottentomatoes [EB/OL]. [2013-05-20]. <http://www.rottentomatoes.com/>.
- [11] BOBADILLA J, ORTEGA J, HEMANDO A, et al. A collaborative filtering approach to mitigate the new user cold start problem [J]. Knowledge-Based Systems, 2012, 26: 225-238.
- [12] AHN H J. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem [J]. Information Sciences, 2008, 178(1/2): 37-51.