

机器学习： 因子分解机模型与推荐系统

燕彩蓉 潘 乔 编著



科学出版社

机器学习：因子分解机模型 与推荐系统

燕彩蓉 潘 乔 编著

科学出版社

北 京

内 容 简 介

因子分解机模型因为能够有效解决高维数据特征组合的稀疏问题且具有较高的预测准确度和计算效率,在广告点击率预测和推荐系统领域被广泛研究和应用。本书对因子分解机模型及其相关模型的研究进展进行综述,阐述该模型的灵活性和普适性,对模型中有待深入研究的难点、热点及发展趋势进行展望。结合研究成果,进一步对该模型进行扩展,并将此扩展后的模型应用于时尚电商领域的推荐任务。提出大数据环境下时尚电商推荐系统框架、研究内容、关键问题,以及可以采用的相关技术,最后通过一个实例验证方案的可行性。

本书适合将要或正在从事数据挖掘、机器学习、推荐系统、时尚电商相关研究的科研人员参考。其中关于因子分解机模型的综述,以及对其改进工作,可以帮助科研人员拓宽研究思路;关于时尚电商领域的推荐系统研究及应用,可以促进机器学习方法在推荐系统中的应用。

图书在版编目(CIP)数据

机器学习:因子分解机模型与推荐系统/燕彩蓉,潘乔编著. —北京:科学出版社,2019.2

ISBN 978-7-03-060145-2

I. ①机… II. ①燕… ②潘… III. ①机器学习—研究 IV. ①TP181

中国版本图书馆CIP数据核字(2018)第287380号

责任编辑:余 丁 / 责任校对:张凤琴

责任印制:吴兆东 / 封面设计:蓝 正

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

北京中石油彩色印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2019年2月第 一 版 开本:720×1000 1/16

2019年2月第一次印刷 印张:7 3/4

字数:146 000

定价:58.00 元

(如有印装质量问题,我社负责调换)

前 言

随着移动互联网、云计算和大数据等技术的发展与应用，推荐已经成为我们生活中的一部分，听音乐、看电影、读书、打游戏、订外卖、旅行等等，推荐都时时伴随着我们。在未来生活中，推荐会变得越来越流行，我们也将越来越依赖于它。好的推荐能够知用户之所需，想用户之所想，帮助用户从海量信息中选择最适合的部分。

因子分解机(factorization machine, FM)模型因为能够有效解决高维数据特征组合的稀疏问题且具有较高的预测准确度和计算效率，在广告点击率预测和推荐系统领域被广泛研究和应用，对 FM 及其相关模型的研究进展进行综述有利于促进该模型的进一步改进和应用。本书通过比较 FM 模型与其他相关模型之间的关联关系，阐述 FM 模型的灵活性和普适性；从特征的高阶交互、特征的场交互、特征的分层交互，以及基于特征工程的特征提取、合并、智能选择和提升等角度总结模型在宽度扩展方面的方法、策略和关键技术；比较和分析 FM 模型与其他模型的集成方式和特点，尤其是与深度学习模型的集成，为传统模型的深度扩展提供思路；同时对 FM 模型的不同优化学习方法和基于不同分布式计算框架的实现进行概括、比较和分析；对 FM 模型中有待深入研究的难点、热点及发展趋势进行展望。

针对 FM 模型对所有特征进行交互造成不必要计算资源浪费的问题，以及噪声可能影响预测精度的问题，我们从特征工程和算法两方面对模型提出改进。首先是对关键属性进行属性提升，运用特征工程技术将因子选择智能地嵌入到算法求解过程中，建立智能化场感知分解机(intelligent FFM, iFFM)模型。iFFM 模型在提高准确度的同时，在一定程度上节省模型的训练时间。

接着，对 iFFM 模型进一步优化，建立广义场感知分解机(generalized FFM, GFFM)模型，即采用多文件有选择地对不同特征编码分开存储替代单个文件存储，降低模型时间复杂度。同时将特征细分为随时间变化的动态特征和保持稳定的静态特征，在建模时结合动态特征的时序变化如物品流行趋势和用户行为偏好，并基于时间窗口建立精准动态模型。

而且，将 GFFM 模型与稠密网络 DenseNet 通过不同方式进行融合，综合利用传统机器学习对低秩特征的快速学习能力以及神经网络对高维特征的提取

能力，进一步提高模型准确度，通过实验分析不同融合方式的优劣，为推荐领域未来的发展方向提供一些数据支撑。

每次在我所培养的研究生的毕业论文中看到致谢导师的话都非常感动，本书内容涉及两届研究生积累下来的成果，这提供了本书的写作素材，在这里表示特别感谢，他们是张青龙、赵雪、黄颜、徐淑华。感谢黄永锋老师对本书内容提出宝贵建议。特别感谢东华大学计算机科学与技术学院为我提供了良好的科研环境，令我心无旁骛，专心致志于科研和教学工作。

“不积跬步，无以至千里；不积小流，无以成江海。”机器学习发展迅速，希望我们的工作能够为相关领域的研究者提供些微帮助。笔者自认才疏学浅，对机器学习领域略知皮毛，更兼时间和精力有限，书中不足之处在所难免，恳请读者不吝赐教，将不胜感激。

燕彩蓉

2018年9月

目 录

前言

第 1 章 绪论	1
1.1 预测和推荐问题描述	1
1.2 研究意义	3
1.3 国内外研究现状及发展动态	4
1.3.1 数据稀疏性和冷启动问题	5
1.3.2 用户偏好和物品流行度动态建模	6
1.3.3 大数据处理和模型的扩展性	7
1.3.4 多样性和准确性平衡问题	8
1.4 本书组织结构	9
第 2 章 FM 模型及其扩展	12
2.1 逻辑回归模型	12
2.2 基于因子分解的多项式回归模型	13
2.3 FM 模型	14
2.4 FM 模型与矩阵分解模型的转化	15
2.4.1 矩阵分解模型	15
2.4.2 FM 模型转化为矩阵分解模型	16
2.5 FM 模型的高阶扩展	17
2.6 FM 模型的场交互扩展	18
2.7 FM 模型的层次交互扩展	19
2.8 FM 模型与其他模型的集成	20
2.9 本章小结	22
第 3 章 特征工程及其对 FM 模型的影响	23
3.1 属性、特征、特征向量和数据集	23
3.2 特征工程	24
3.3 特征的来源	25
3.4 FM 模型相关研究中的特征工程	27
3.5 FM 模型的应用领域	30

3.6	本章小结	32
第4章	模型训练方法	33
4.1	预测和推荐模型的目标优化	33
4.2	模型训练方式	34
4.2.1	拟牛顿法	34
4.2.2	SGD 系列算法	35
4.2.3	Gibbs 采样算法	37
4.3	激活函数	38
4.4	过拟合问题	39
4.4.1	正则化方式	40
4.4.2	批规范化	41
4.4.3	Dropout 及相关优化方法	42
4.5	本章小结	44
第5章	智能化场感知分解机	45
5.1	算法改进思路	45
5.2	iFFM 模型	46
5.3	多样性处理	48
5.3.1	热扩散算法	48
5.3.2	两个模型的集成	50
5.4	实验结果与分析	50
5.4.1	实验环境	50
5.4.2	实验结果	51
5.5	本章小结	54
第6章	广义场感知分解机	55
6.1	模型改进思路	55
6.2	时间因子	57
6.3	动态模型构建	58
6.4	GFFM 模型评价	60
6.4.1	实验设置	60
6.4.2	实验结果及分析	62
6.5	本章小结	64
第7章	FM 模型与深度学习模型的集成	65
7.1	FNN 模型	65
7.2	Wide&Deep 模型	66

7.3	Deep&Cross 模型	67
7.4	DeepFM 模型	68
7.5	NFM 与 AFM 模型	68
7.6	宽度和深度学习模型集成方式分析	69
7.7	本章小结	70
第 8 章	基于稠密网络的广义场感知分解机	71
8.1	ResNet 和 DenseNet	71
8.2	DGFFM 模型	73
8.2.1	Wide&Deep 结构	73
8.2.2	FNN 结构	74
8.3	DGFFM 模型评价	75
8.3.1	实验设置	75
8.3.2	实验结果及分析	75
8.4	本章小结	76
第 9 章	FM 模型实现库及并行化处理	77
9.1	libFM	77
9.1.1	libFM 中核心类之间的关系	77
9.1.2	fm_learn 类代码解析	78
9.2	FM 的其他实现库	83
9.2.1	libFFM	83
9.2.2	fastFM	83
9.3	FM 模型的其他优化方法	83
9.4	FM 模型的并行实现	84
9.5	本章小结	87
第 10 章	时尚电商领域的推荐系统研究	88
10.1	深度学习为时尚推荐研究带来新思路	88
10.2	大数据环境下时尚电商推荐系统框架及面临的问题	89
10.3	融合视觉特征的推荐系统研究内容	90
10.3.1	融合非视觉属性与视觉特征及其时空动态性的推荐模型研究	91
10.3.2	面向大规模数据的并行化模型训练算法与技术研究	92
10.3.3	在线推荐引擎研究	92
10.4	关键问题	93
10.4.1	基于视觉特征的时空动态建模	93
10.4.2	模型训练的优化问题	93

10.4.3	推荐引擎的实时处理	93
10.5	相关技术	94
10.5.1	基于深度卷积神经网络的视觉特征提取方法	94
10.5.2	基于 iFM 模型的非视觉属性建模	95
10.5.3	基于 VBPR 的视觉特征建模	96
10.5.4	基于马尔可夫链的时间序列预测	96
10.5.5	基于 SGD 算法的模型训练方法	97
10.5.6	基于 TensorFlow 的并行算法实现	98
10.6	本章小结	98
第 11 章	一个 N-阶段购买决策模型	100
11.1	研究背景	100
11.2	特征处理与 Wide&Deep plus 框架	101
11.3	NSPD 模型及其优化	102
11.4	实验与结果评价	105
11.4.1	数据集	105
11.4.2	评价指标	106
11.4.3	算法实现	106
11.4.4	实验结果	106
11.5	本章小结	108
参考文献		109

第 1 章 绪 论

云计算、移动互联网和社会媒体等技术的迅猛发展,使得网络空间中蕴含的信息量呈指数级增长。作为缓解信息过载问题的有效手段,推荐系统得到学术界和工业界的广泛关注,相关研究成果已经融入我们日常生活的各项个性化服务中,如在线电商、信息检索和移动应用。推荐系统有两大任务:评分预测和 top- k 推荐,前者用于预测用户对于未接触事物的评分,后者为用户提供个性化的推荐列表。面对多源、异构、动态的大数据,传统推荐模型或算法在准确性、扩展性、实时性方面都面临着挑战。

1.1 预测和推荐问题描述

预测和推荐是两类非常相关的任务,预测任务通常包括三种:分类、回归和排序^[1]。两者之间的关系体现为:预测的结果可为推荐提供服务,推荐需要以预测为基础。

预测是指通过对训练集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ 进行学习,建立一个从输入空间 X 到输出空间 Y 的映射 $f: X \rightarrow Y$ 。若输出空间 Y 离散,那么这类学习任务被称为分类(classification),若分类中只涉及两个类别,即 Y 中只包含两个值,则称为二分类(binary classification);涉及多个类别则称为多分类(multi-class classification)。若输出空间 Y 连续,那么这类任务称为回归(regression)。对于二分类任务, $Y = \{-1, +1\}$ 或 $\{0, 1\}$;对于多分类任务, $Y = \{0, 1, \dots, n\}$, 其中 n 为类别的数量;对于回归任务, Y 通常为实数集 $[k_1, k_2]$, 其中 k_1 为回归下限, k_2 为回归上限,一般采用相关函数(如 clip)将回归控制在一定范围内。预测的一个典型应用就是点击率(click-through rate, CTR)预测,预测结果只有两个,点击或者不点击,可以将其转化为二分类问题。

点击率是互联网广告常用的术语,指网络广告(图片广告、文字广告、关键词广告、排名广告、视频广告等)的点击到达率,即该广告的实际点击次数除以广告的展现量。点击率预测是互联网主流应用(广告、推荐、搜索等)的核心算法问题,谷歌(Google)和脸书(Facebook)等业界巨头对这个问题一直进行着持续的投入和研究。点击率预测是互联网计算广告中的关键技术环节,其预测的准确性

直接影响公司的广告收入。大数据时代，广告领域的点击率预测问题面临着超高维离散特征空间中模式发现的挑战，即如何拟合现有数据的规律，同时又具备推广性。

推荐是指通过分析和挖掘用户(user)与物品(item)之间的二元关系及相关属性，帮助用户从海量数据中发现其感兴趣的有形或无形的物品(如信息、服务、商品等)，然后生成个性化推荐列表^[2,3]。传统的属性主要包括用户和物品的固有特征、显式的用户评分，以及隐式的用户反馈。隐式反馈普遍存在于互联网中，如用户在观看电影、收听音乐、浏览或购买商品等时留下的行为痕迹。近几年来利用社交网络和移动位置信息进行协同过滤推荐成为研究热点^[4,5]。由于互联网中用户反馈多以隐式反馈的形式存在，且用户平均浏览商品数一般远小于商品总数，因此设计矩阵(用户物品矩阵)通常是稀疏的。而这些稀疏的数据场景为预测和推荐的研究带来新的挑战。推荐系统和点击率预测都可以看作是分类问题在真实场景中的应用。

预测和推荐系统通常由三部分组成：特征提取及处理、模型建立与训练、在线服务，如图 1.1 所示，所有工作都构建于特征之上。在此框架中，首先是特征提取及处理，不同的特征经过处理后进行融合；然后输入到模型中，对模型进行训练和验证，验证通过的模型将为在线预测和推荐提供依据；最后用户的反馈将重新加入特征中，为模型的改进和再次训练提供数据。考虑到推荐任务中特征并不总是连续的，多数情况是类别值，所以将这些特征数字化将更适合于模型训练。为了应对一些无序的类别特征，可选用独热编码(one-hot encoding)方式对其进行编码。

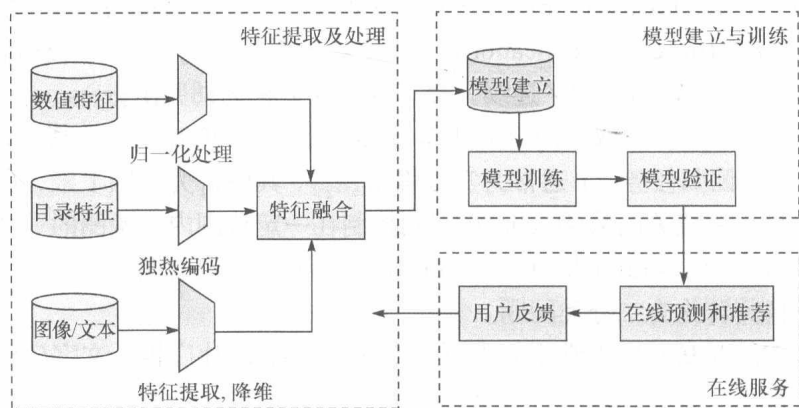


图 1.1 预测和推荐系统框架

独热编码又称一位有效编码，是广泛使用的一种特征表示方法。具体操作为：

使用 N 位状态寄存器对 N 个状态进行编码, 每个状态都有其独立的寄存器位, 并且在任意时候, 只有一位有效。独热编码解决了分类器不易处理属性数据的问题, 在一定程度上扩充了特征。独热编码可将分类特征转化为数值型特征, 从而方便进行模型预测。假设某电子商务网站中耳机的品牌包括 Panasonic、Sennheiser、Sony、Sound Intone、Bose、Beats、Audio-Technica 和 Ausdom 共 8 种, 采用一个长度为 8 的向量对耳机品牌进行独热编码, 特征 Panasonic 可表示为 $(1,0,0,0,0,0,0,0)$, 特征 Sennheiser 可以表示为 $(0,1,0,0,0,0,0,0)$, 以此类推。图 1.2 描述的是一个推荐问题, 特征用 x 表示, 每个 x_i 表示一个特征向量, 对应的目标值为 y_i 。图 1.2 中相关数据来源于 Rendle 的论文^[6], 第 1 组表示用户特征, 采用独热编码; 第 2 组表示物品特征, 采用独热编码; 第 3 组表示同一用户对其他物品的评分; 第 4 组表示时间特征; 第 5 组表示本用户最后评分的物品, 采用独热编码。通过这些数据对模型进行训练, 验证好的模型可以提供在线预测和推荐。可以看出, 这些特征所构成的矩阵存在着大量的数据冗余, 而且数据也非常稀疏(0 元素非常多)。

		特征向量 x															目标 y					
x_1	1	0	0	...	1	0	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	0	...	5	y_1
x_2	1	0	0	...	0	1	0	0	...	0.3	0.3	0.3	0	...	14	1	0	0	0	...	3	y_2
x_3	1	0	0	...	0	0	1	0	...	0.3	0.3	0.3	0	...	16	0	1	0	0	...	1	y_3
x_4	0	1	0	...	0	0	1	0	...	0	0	0.5	0.5	...	5	0	0	0	0	...	4	y_4
x_5	0	1	0	...	0	0	0	1	...	0	0	0.5	0.5	...	8	0	0	1	0	...	5	y_5
x_6	0	0	1	...	1	0	0	0	...	0.5	0	0.5	0	...	9	0	0	0	0	...	1	y_6
x_7	0	0	1	...	0	0	1	0	...	0.5	0	0.5	0	...	12	1	0	0	0	...	5	y_7
		A	B	C	D	TI	NH	SW	ST	...	TI	NH	SW	ST	...	TI	NH	SW	ST	...		
		用户				电影					其他电影排名					时间	其他排名的电影					

图 1.2 特征向量示例

1.2 研究意义

随着技术进步和移动物联网的发展, 数据信息呈爆炸式增长。根据智研咨询发布的《2016—2022 年中国大数据行业深度分析及投资战略咨询报告》^①, 2015 年, 全球数据存储量已经达到 8.61ZB。假设地球上每个人(联合国估计 2017 年地球人口总数为 75 亿)每分钟发 3 条推特信息(每条 140 字节), 大约需要连发

① <http://www.chyxx.com/research/201608/439946.html>

6100年。在中国,2017年“双十一”一天时间,阿里平台总成交额已经达到1682亿人民币,产生物流订单8.12亿个,创造了新的“中国速度”及“世界高度”,由此而来的数据相当可观。数据量的快速增长造成了大量的信息过载,如何有效过滤信息,找到最有价值的部分成了企业乃至国家发展的重要战略目标。

作为一种有效的信息过滤手段,推荐系统是当前解决信息过载问题及实现个性化信息服务的有效方法之一^[4]。目前主流的推荐系统应用于很多领域,包括:电子商务,如Netflix、Amazon、eBay、阿里巴巴和豆瓣等^[7];信息检索,如iGoogle、MyYahoo、GroupLens和百度等^[8];移动应用,如Daily Learner和Appjoy等^[9];生活服务,如美团和新浪微博等^[10]。在计算广告领域,常用的过滤手段是点击率和转化率(conversion rate, CVR)预测。准确地估计点击率和转化率对于提高流量的价值,增加广告收入有重要的指导作用。常用的预估点击率和转化率的方法有Google的在线学习算法FTRL-Proximal^[11]、Facebook的GBDT(gradient boosting decision tree)+LR(logistic regression)^[12]、百度最近使用的深度神经网络(dense neural network, DNN)等。

推荐算法一般采用的方式有top- k 推荐和评分预测两种,这两种方式可以归结为一类预测问题。top- k 可预测用户最感兴趣的 k 个物品,评分预测的结果如果以概率的形式输出,则可以认为是用户对物品评分的概率(如5分的概率为0.4,3分的概率为0.6,其余为0),根据概率排序,选择概率最高的 k 个物品,即为top- k 推荐。由于目前多数网站的用户评分形式都采用整分制(单个用户不允许评分为小数),因此评分预测可以转化为多分类问题。点击率预测的结果只有点击或不点击两个,可以转化为二分类问题。因此,推荐系统和点击率预测都可以看作是分类问题在真实场景中的应用。

1.3 国内外研究现状及发展动态

推荐系统在国内外赢得了广泛关注,许多大学和研究机构对此领域展开了深入研究。国外的研究机构主要有Microsoft研究院、Yahoo研究院、Google研究院、加利福尼亚大学等;国内的研究机构主要有中国科学院、清华大学、北京大学、上海交通大学、武汉大学、北京邮电大学、中国人民大学等。各大电子商务网站,如Netflix、Amazon、Zalando、淘宝等,利用推荐系统为用户提供更多的服务并以此获利。国际重要学术会议近几年的主题中均包含推荐系统研究,如数据挖掘领域的SIGIR、SIGKDD、ICDM等。ACM推荐系统年会(ACM Conference on Recommender Systems, RecSys)是一个顶级的国际论坛,每年为推

荐系统领域展示新的方法、技术和框架。2016年第1届推荐系统中的深度学习会议(Workshop on Deep Learning for Recommender Systems, DLRS)与ACM推荐系统年会同时开展,因为推荐系统领域专家认为:近几年深度学习在计算机视觉、自然语言处理和语音识别等领域已经取得较大的成功,然而在推荐系统领域开展的工作还较少,深度学习已经成为推荐系统领域研究的下一个重要热点。下面从几个方面具体阐述其发展动态。

1.3.1 数据稀疏性和冷启动问题

现实生活中,数据矩阵往往极其稀疏,以电影《魔兽》为例,中国票房为14.7亿元,观影人次近4000万,豆瓣评分人次为181 431,评分密度仅为0.45%。为了缓解数据稀疏问题,学者提出了社会化推荐(social recommendation)方法^[13,14]。为了提取更优的潜在特征向量,文献[15]将用户的各种社会网络关系融合到矩阵的优化分解过程中,提出社会化的矩阵分解(social matrix factorization)。文献[16]指出用户对物品的评分信息与用户之间的社会关系信息往往来源于不同的数据源,社会化推荐具有一定的局限性。作者在贝叶斯概率矩阵分解(Bayesian probabilistic matrix factorization, BPMF)^[17]的基础上,使用更加稀疏的广义高斯分布取代原始的高斯分布初始化数据,取得了较好的效果。类似的方法还有Glorot initialization^[18]和He initialization^[19]。Glorot initialization的基本思想是保持输入和输出的方差一致,这样就避免了所有输出值都趋向于0,在激活函数为tanh时,使用神经网络训练,输出值在很多层之后仍能保持着良好的分布。He initialization的核心思想是在使用修正线性单元(rectified linear unit, ReLU)的神经网络中,每一层仅有一半的神经元被激活,另一半为0,所以要保持方差不变,需要在Glorot的基础上除以2才可。

推荐系统的另一个问题是用户和物品的冷启动问题。对于新用户,一般有两条途径来缓解冷启动问题:第一种,选择从第三方渠道,如QQ或微信,登录后导入其信息(如果有);第二种,引导用户把自己的一些属性表达出来,如回答一系列问题,选择喜欢的物品,添加属性标签等。这两种方式有时也会同时使用。若无法有效获取用户相关信息,一般常用的手段是推荐热门物品。对于新物品,可以利用它们的一些属性,如电子商务时尚推荐中物品的颜色、样式和品牌等,计算与现有物品的相似性,进而推荐给相关的用户。对于完全新颖的物品,一个比较好的推荐手段是利用品牌,比如“索粉”如果在购买力允许的情况下,一般会比较愿意尝试索尼的新产品,这种方式有时也被称为“粉丝”营销。

从模型的角度分析,则需要算法能够捕捉到用户或者物品更多的特征。常用的逻辑回归等线性方法能够很好地利用属性的单独特征或者一阶特征(raw

feature)。交叉特征(cross feature)通常依赖于人工设计，比如电影票容易在周五晚上被用户购买，就需要构建电影 ID 和时间之间的联系。然而，大量的特征隐藏在数据中，很难被专家识别和设计，这就需要设计算法来解决。Poly-2(degree-2 polynomial mappings)^[20,21]直接对二阶特征组合建模来学习权重。FM 模型^[22]为每个单独特征构建隐式向量，并通过隐式向量之间的内积(inner product)来建立两个特征的组合关系，从而实现对二阶特征组合的自动学习。FM 模型因综合了矩阵分解(matrix factorization, MF)和支持向量机(support vector machine, SVM)之间的优点，能够有效解决高维数据特征组合的稀疏问题，且具有线性时间复杂度，近几年被广泛研究和应用^[6]。

Juan 及其比赛团队借鉴 Jahrer 论文^[23]中的场(field)概念，提出了场感知分解机(field-aware factorization machine, FFM)模型^[24]。FFM 模型引入场概念，把相同性质的特征归于同一个场。自提出以来，FFM 模型表现突出，分别在 Criteo^①和 Avazu^②举办的点击率预测竞赛中夺得冠军。文献[25]指出 FFM 模型不仅能够赢得比赛，而且在真实预测系统中具有很高的价值。在中国，美团点评技术团队在搭建需求方平台(demand side platform, DSP)的过程中，探索并使用了 FM 和 FFM 模型进行点击率和转化率预估，取得了不错的效果。然而，和 FM 模型一样，FFM 模型对所有特征变量进行交互建模，通过共享特定特征隐式向量来计算因式分解参数。

真实环境中，特征变量通常很多，而且并非所有交互都有效，为了提高计算效率，本书对 FFM 模型进行改进，提出了智能化场感知分解机(intelligent FFM, iFFM)模型^[26]。iFFM 模型对关键属性进行提升，运用特征工程技术将因子选择智能地嵌入到算法求解过程中，并综合利用 Gibbs 采样(Gibbs sampling)和随机梯度下降(stochastic gradient descent,SGD)算法来提高推荐准确度。对于高维特征的提取，DNN 提供了一个很好的解决方案。比如，对于图片常采用卷积神经网络(convolutional neural network, CNN)来提取特征，然后对特征进行降维处理。也有将低秩特征(一阶特征和二阶特征)与高维特征结合起来一起建模的方法，如 Wide&Deep, DeepFM 等。

1.3.2 用户偏好和物品流行度动态建模

对用户偏好和物品流行趋势建模也是推荐系统面临的挑战之一，尤其是在时尚推荐领域。比如，Google 将流行趋势划分为 6 类：持续上升(sustained risers)、

① <https://www.kaggle.com/c/criteo-display-ad-challenge>

② <https://www.kaggle.com/c/avazu-ctr-prediction>

季节上升(seasonal risers)、突然上升(rising stars)、持续下降(sustained decliners)、季节下降(seasonal decliners)、突然下降(falling stars)等,如图 1.3 所示^①。通过划分各流行趋势,对各个趋势变化分别进行分析。文献[27]提出了时尚 DNA(fashion DNA)的概念,使用 DNN 对时尚物品进行特征提取,结合其他特征,为每个物品生成唯一的标签,类似于人类的 DNA,这个标签值能够唯一地确定一个物品。文献[28]将时间动态因素应用到奇异值分解(singular value decomposition, SVD)中,分别对用户偏置和物品偏置进行动态建模,然后与 SVD 模型进行结合。文献[29]对视觉特征(visual feature)和非视觉特征(non-visual feature)进行轻量级时间建模,然后进行特征的融合。由于用户行为变化趋势和物品变化趋势有很大的不同,比如用户行为在工作日和周末可能有很大不同(工作日忙于上班,周末可能更注重休闲娱乐),而物品短期内流行趋势可能没有明显变化,因此将用户和物品按照同样的方式进行时间建模并不是很合理。

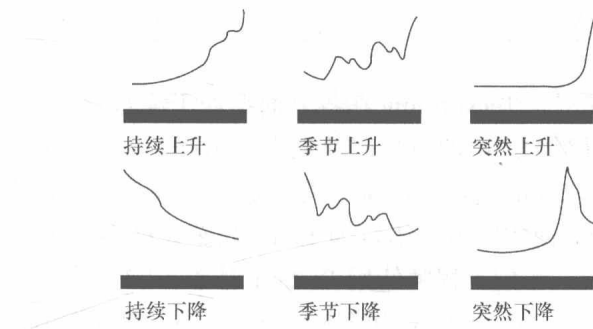


图 1.3 时尚流行趋势变化

和多数领域一样,推荐系统和点击率预测同样存在大数据和模型扩展性问题。大数据应用方面,目前主流的解决方案是使用分布式框架和并行计算平台(如 NVIDIA® CUDA®等);硬件方面,则使用 GPU、TPU 等进行加速。在线服务平台使用较多的有 Amazon 的 AWS、Google 的 Cloud 等。中国境内常用的云平台有阿里云和腾讯云。若条件允许,中国境内的高校和研究机构则更倾向于搭建自己的集群服务平台。下一节将介绍几种在深度学习方向常用的支持 CUDA 的框架。

1.3.3 大数据处理和模型的扩展性

推荐系统是大数据时代下应运而生的产物。随着数据规模增大,数据类型增

^① <http://www.yesstyle.com/blog/tag/google-trend-2016/>

多，数据来源越来越复杂，推荐处理更加依赖于大数据处理框架。

目前，深度学习领域持续火热，各种开源框架层出不穷，常用的有 TensorFlow、Caffe、MXNet、Keras、Theano、PyTorch 等。本书对各开源框架在 GitHub 上的数据信息从 Watch 数量、Star 数量、Fork 数量、Contributor 数量 4 个方面进行简单统计(统计时间为 2018 年 8 月 1 日)，如表 1.1 所示。

表 1.1 GitHub 上各个开源框架的数据统计

框架	机构	支持语言	Watch	Star	Fork	Contributor
TensorFlow	Google	Python/C++/...	8 222	106 370	36 827	1 578
Caffe	BVLC	Python/C++	2 220	25 047	15 324	269
MXNet	Apache	Python/C++/...	1 161	14 740	5 401	569
Keras	Google	Python	1 784	32 083	12 029	703
Theano	LISA Lab	Python	587	8 392	2 226	326
PyTorch	Facebook	Python	925	17 499	4 096	723

通过表 1.1 可以看出，TensorFlow 在各方面都处于绝对领先的地位，这得益于 Google 在深度学习领域杰出的贡献以及强大的感召力。上述几种框架都支持 Python 语言，可见 Python 语言在深度学习领域的重要地位。值得一提的是，PyTorch 的前身 Torch7 使用的编程语言为 Lua。Lua 太过小众导致 Torch7 发展前景堪忧，于是 Facebook 的工程师使用 Python 对 Torch7 进行重写，这才有了今天熟悉的 PyTorch。Theano 最初是为学术研究设计的，易于与其他深度学习库结合，灵活性很强，使用比较麻烦。目前，Theano 已经停止更新。另外，Keras 也是一种常用的深度学习框架，同样来自 Google。在构建神经网络方面，相较于 TensorFlow，Keras 具有快速简单等特点，但 Keras 扩展能力差，实现其他机器学习算法比较困难，而现实中模型的构建往往是极其复杂的，很难用 Keras 实现。因此，Keras 通常只用于学术研究。

从机器学习库的实现角度总结如下：TensorFlow 是相对高阶的机器学习库，用户拥有更高的对网络结构设计的自主权；Caffe 的核心概念是 Layer，在计算机视觉领域的应用较多，常用来做人脸识别、图片分类、位置检测、目标追踪等；Keras 是一个非常高层的库，可以工作在 Theano 和 TensorFlow 之上。另外，Keras 强调极简主义，用户只需几行代码就能构建一个神经网络。

1.3.4 多样性和准确性平衡问题

对于点击率预测，一般情况下每个用户浏览的界面只能显示一个广告，所以不需要考虑多样性。对于推荐方向，每个页面可以显示多个推荐结果，这时推荐