

华侨大学本科毕业论文(设计)开题报告

学院： 计算机科学与技术

专业班级： 2014计算机科学与技术

姓名	潘傲寒	学号	1425111035	指导教师	王成	职称 学历	副教授 博士研究生
课题名称		基于深度学习的协同过滤推荐算法研究					
设计（论文）类型 （划√）	工程设计	应用研究	开发研究	基础研究	其它		
				√			
<p>1、本课题的研究目的和意义：</p> <p>目的： 将深度学习应用到协同过滤推荐算法中，研究深度学习是否能够显著改善传统的推荐系统，并使用主动学习的原则来指导新知识的获取。</p> <p>意义： 随着互联网中数字信息数量的增长，在线商店、在线音乐、视频和图片库、搜索引擎和推荐系统已成为人们快速寻找信息的主要方式。 近年来，推荐系统被证明是一种解决信息过载问题的有效工具。从根本上来讲，推荐系统是通过为用户指引改用户不熟悉的新物品来解决信息过载现象的，这些新物品或许与该用户当前的需求相关。 推荐系统的协同过滤算法已经得到了人们的普遍关注并取得了很大的进展。协同过滤算法基于用户对商品的评分或其它行为模式来为用户提供个性化的推荐，而不需要了解用户或者商品的大量信息。 深度学习的发展在语音识别、图像处理和自然语言处理领域取得了很多成果。此外，深度学习在推荐系统和信息检索领域里也得到应用。协同过滤推荐算法是推荐系统的核心。将深度学习应用在协同过滤领域，并研究深度学习是否能显著改善传统的推荐系统，有很重要的意义。</p>							
<p>2、文献综述（国内外研究情况及其发展）：</p> <p>我们的日常需求，从商品、书籍、新文章、歌、电影、研究文件和其他基础性事物，其数量和种类填满了多个数据仓库和数据库。为此，智能推荐系统和强大的搜索引擎为用户提供援助之手。此类系统的流行程度和有用性在于它们能够便捷地显示几乎无限的物品信息 [1]。因此 Amazon、Netflix 等推荐系统尝试了解用户兴趣，并把用户感兴趣的商品告知他们。尽管这些系统由于使用场景而各不相同，但其寻找用户感兴趣商品的核心机制都是用户兴趣与商品匹配的机制 [2]。</p> <p>国内外研究深度学习应用至协同过滤推荐系统的成果中，大多数方法都在尝试使用一些深度神经网络代替矩阵分解。</p> <p>[3] 首次提出了将深度学习特征融合到 CF 模型（如矩阵分解）的框架。作者通过将该模型与几种使用矩阵分解的协同过滤方法进行比较，展现了其在 4 个现实世界数据集（MovieLens-100k、MovieLeans-1M、Book-Crossing 和 Advertising 数据集）上的性能提升。</p> <p>对于基于会话的推荐系统（session recommendation），一些研究使用循环神经网络来提升推荐系统的性能。[4] 首先使用 RNN 提升基于短会话数据的推荐系统。该论文的作者表明传统的矩阵分解方法不适合基于会话的推荐系统。研究人员在电子商务点击流数据和类似 YouTube 的 OTT 视频服务数据集上进行了实验，结果表明使用 RNN 要比 MF 方法更有优势。</p> <p>此外，[5] 使用图像和文本等特征以加强基于 RNN 的会话推荐系统。论文作者引入了并行 RNN 的概念，并用来对特征属性（如文本、图像、ID 等）进行建模。与简单的 RNN 和 item-kNN 相比，经过特征增强的 RNN 在类似 YouTube 的视频服务数据集上有显著的性能提升。</p> <p>而为推荐系统增加主动学习功能，能够帮助用户更好地意识到自己的好恶，同时可以为系统提供新信息以便后续的推荐。</p> <p>基于对话的主动学习[6, 7, 8]，它通过问题的迭代引出用户响应来减少在快速发现用户寻找内容时的搜索空间。</p> <p>基于不确定性的主动学习尝试获取训练点来减少某些方面的不确定性，例如，与输出值相关[9]、模型参数[10]和决策边界等。</p>							

3、本课题的主要研究内容（提纲）和成果形式：

1. 将协同过滤变治转化为序列预测问题
2. RNN 若干种变体以及在协同数据上的最优变体
3. 对比传统协同过滤方法，探索深度学习在协同过滤上的优势
4. 利用主动学习，使得深度学习协同过滤算法能够使用少量样本，迅速建立准确的模型。

成果形式：

论文 1 篇

4、拟解决的关键问题：

1. 协同过滤如何转化为序列问题
2. RNN 及其变体如何与协同系统结合
3. 将主动学习应用于推荐系统

5、研究思路、方法和步骤：

课题将探索若干个深度学习模型，并与传统方法进行对比。

一、RNN 及其变种

RNN: 与前馈型网络不同，“递归神经网络”(recurrent neural networks)允许网络中出现环形结构，从而可让一些神经元的输出反馈回来作为输入信号。这样的结构与信息反馈过程，使得网络在 t 时刻的输出状态不仅与 t 时刻的输入有关，还与 $t-1$ 时刻的网络状态有关，从而能处理与时间有关的动态变化。

LSTM: [11] 循环神经网络(RNN)在工作时一个重要的优点在于，其能够在输入和输出序列之间的映射过程中利用上下文相关信息。然而不幸的是，标准的循环神经网络(RNN)能够存取的上下文信息范围很有限。这个问题就使得隐含层的输入对于网络输出的影响随着网络环路的不断递归而衰退。因此，为了解决这个问题，长短时记忆(LSTM)结构诞生了。LSTM 能够更好地对序列数据建模。

GRU: [12]GRU 作为 LSTM 的一种变体，将忘记门和输入门合成了一个单一的更新门。同样还混合了细胞状态和隐藏状态，加诸其他一些改动。最终的模型比标准的 LSTM 模型要简单，训练速度比 LSTM 快，是非常流行的变体。

SRU: [13]SRU 简化了状态计算的过程，从而展现出了与 CNN、注意力和前馈网络相同的并行性。极大加快了训练速度，是 LSTM 训练速度的 5 倍以上。

二、对比方法

有许多常用的非深度学习协同过滤方法，本课题将用同样的评测指标对几种协同过滤方法和深度学习方法进行对比。

User-Based Nearest Neighbors: 基于用户的协同过滤方法是较为常用的协同过滤方法之一，课题会在相同的数据集上进行验证比较。

BPR-MF: 是一个很先进的矩阵分解方法，它基于 BPR ranking 损失。

Markov Chain: 马尔可夫链，满足马尔可夫性质的随机过程，适用于时序建模。在协同过滤变治为时序问题后，马尔可夫链用来与深度学习模型在时序建模中比较。

三、评价指标

sps.: 短期预测成功率(The Short-term Prediction Success)描述了一个方法预测(推荐)下一个物品的能力。

Recall: 召回率是一个常用于 top-N 推荐的评价指标，它描述一个方法做长期预测(推荐)物品的能力。

User coverage: 用户覆盖率指用户收到至少一个正确推荐的比率。平均召回率隐藏了用户被成功推荐的分布，因为一个高用户覆盖率不代表许多用户收到

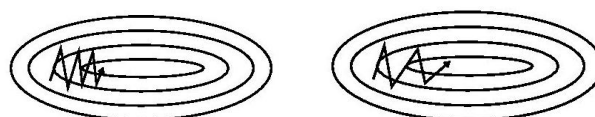
Item coverage: 物品覆盖率指多少物品被正确推荐。它度量一个方法的多样性。

四、优化算法

现代神经网络很少使用随机梯度下降法训练，有许多优化算法被发明出来加速训练过程。

Momentum: SGD 在遇到沟壑时容易陷入震荡，为此，引入动量，加速 SGD 在正确的方向并抑制震荡。

$$m_t = \gamma m_{t-1} + \eta g_t$$

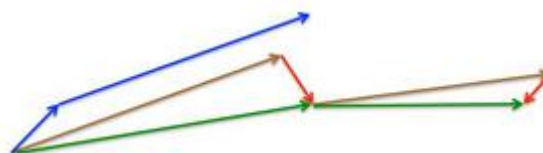


SGD

Momentum

Nesterov Accelerated Gradient: NAG 在 SGD with Momentum 的基础上改进了梯度公式：

$$g_t = \nabla_{\theta} J(\theta - \gamma m_{t-1})$$



SGD with Momentum 计算了当前梯度(短蓝向量)和动量项(长蓝向量)。同时，计算出下一时刻 θ 的近似位置(棕向量)，并根据该未来位置计算梯度(红向量)，然后使用和 SGD with Momentum 中相同的方式计算步长(绿向量)。这种计算梯度的方式可以使算法更好的预测未来，提前调整更新速率。

Adagrad: SGD、SGD-M 和 NAG 均是以相同的学习率去更新 θ 的各个分量。而深度学习模型中往往涉及大量的参数，不同参数的更新频率往往有所区别。对于更新不频繁的参数(典型例子：更新 word embedding 中的低频词)，我们希望单次步长更大，多学习一些知识；对于更新频繁的参数，我们则希望步长较小，使得学习到的参数更稳定，不至于被单个样本影响太多。

Adagrad[4] 算法即可达到此效果。其引入了二阶动量：

$$v_t = \text{diag}(\sum_{i=1}^t g_{i,1}^2, \sum_{i=1}^t g_{i,2}^2, \dots, \sum_{i=1}^t g_{i,d}^2)$$

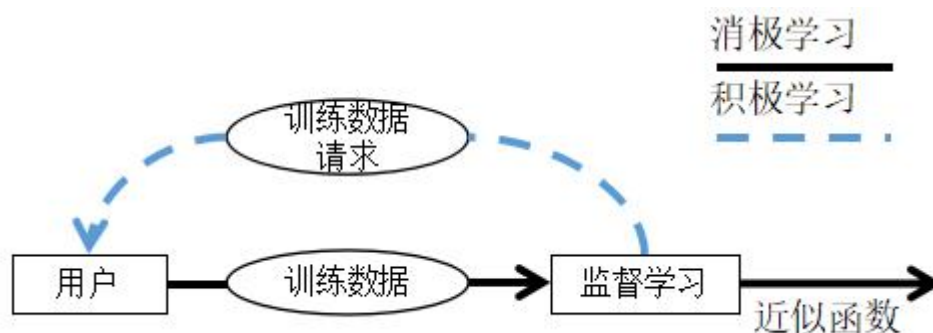
RMSprop: 在 Adagrad 中, V_t 是单调递减的, 使得学习率逐渐递减至 0, 可能导致训练提前结束。为了改进这一缺点, 考虑在计算二阶动量时不累积全部历史梯度, 而只关注最近某一时间窗口内的下降梯度。

$$v_t = \gamma v_{t-1} + (1 - \gamma) \cdot \text{diag}(g_t^2)$$

其二阶动量采用指数移动平均公式计算, 这样即可避免二阶动量持续累积的问题。

Adam: 可以认为是 RMSprop 和 Momentum 的结合。和 RMSprop 对二阶动量使用指数移动平均类似, Adam 中对一阶动量也是用指数移动平均计算。

五、主动学习



传统的通过不断积累训练集的方式被称为被动学习, 它属于一种静态的随机采样。被动学习忽略了学习器本身也有跟环境交流获取数据的能力, 主动学习正是利用了学习器的这种能力, 通过模拟人的学习过程, 选择合适的样本加入训练集。

方法	描述/目标	可能的限制条件
基于不确定性	减少以下内容的不确定性 <ul style="list-style-type: none"> ● 评分评估 ● 决策边界 ● 模型参数 	降低不确定性可能并不能总是提高准确率; 该模型可能只对错误的事情肯定 (如用错了预测方法)
基于误差	通过错误与以下内容之间的关系来减少预测错误 <ul style="list-style-type: none"> ● 输出评估中的变化 ● 测试集错误 ● 参数评估中的变化 ● 参数估计值的方差 	可靠地评估错误的减少是困难并且计算昂贵的
基于组合	基于以下内容之间的共识找出有用的训练集 <ul style="list-style-type: none"> ● 聚类中的模型 ● 多重候选模型 	由于其效用取决于模型/候选的质量且可以是计算昂贵的, 因为它的执行与多重模型/候选有关

6、本课题的进度安排:

2017.12-2018.1 文献资料搜集, 研究深度学习
2018.1-2018.3 做实验, 改进算法
2018.4-2018.5 编写论文
2018.5-2018.6 修改论文, 准备答辩

7、参考文献：

- [1] A. Singhal, “Leveraging open source web resources to improve retrieval of low text content items,” ProQuest Diss. Theses, p. 161, 2014.
- [2] A. Singhal, R. Kasturi, V. Sivakumar, and J. Srivastava, “Leveraging Web intelligence for finding interesting research datasets,” in Proceedings – 2013 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2013, 2013, vol. 1, pp. 321–328.
- [3] S. Li, J. Kawale, and Y. Fu, “Deep Collaborative Filtering via Marginalized Denoising Auto-encoder,” in Proceedings of the 24th ACM International Conference on Information and Knowledge Management – CIKM ’ 15, 2015, pp. 811–820.
- [4] A. K. Balazs Hidasi, “Session-based Recommendation with Recurrent Neural Networks,” ICLR, pp. 1–10, 2016.
- [5] B. Hidasi, M. Quadrana, A. Karatzoglou, and D. Tikk, “Parallel Recurrent Neural Network Architectures for Feature-rich Session-based Recommendations,” in Proceedings of the 10th ACM Conference on Recommender Systems – RecSys ’ 16, 2016, pp. 241–248.
- [6] McGinty, L., Smyth, B.: On The Role of Diversity in Conversation Recommender Systems. Case-Based Reasoning Research and Development pp. 276–290(2003)
- [7] Ricci, F., Nguyen, Q.N.: Acquiring and revising preferences in a critique-based mobile recommender system. IEEE Intelligent Systems 22(3), 22–29(2007). DOI <http://dx.doi.org/10.1109/MIS.2007.43>
- [8] Bridge, D., Ricci, F.: Supporting product selection with query editing recommendations. In: RecSys’ 07: Proceedings of the 2007 ACM conference on Recommender systems, pp. 65–72.
- [9] Kohrs, A., Merialdo, B.: Improving collaborative filtering for new users by smart object selection. In: Proceedings of International Conference on Media Features (ICMF) (2001)
- [10] Hofmann, T.: Collaborative filtering via gaussian probabilistic latent semantic analysis. In: SIGIR’ 03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 259–266.
- [11] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.
- [12] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259, 2014.
- [13] T. Lei, Y. Zhang and Y. Artzi. Training RNNs as Fast as CNNs. arXiv preprint arXiv:1709.02755

8、指导教师意见：

指导教师（签名）：

年 月 日