**基于知识图谱的多视角新闻推送系统设计与实现**

**项目简介：**

目前主流的新闻推送系统已经能够较好地满足用户的主动表达的需求。基于内容及协同过滤并结合用户行为的方法力图使推送效果呈现“千人千面”。虽然完全迎合用户行为习惯进行精准推送能够极大地提高准确性，但往往会局限化用户获取信息的范围。本项目的基本思想是不仅仅为用户推送所需的已知信息，关键在于推送未知的新知识，拓宽视野，将推送过程变成发现之旅，从而提升用户体验。本项目利用知识图谱设计多视角注意力模型，综合新闻文本的准确性、时效性和全面性，在神经网络中进行大规模数据训练。同时，避免仅以准确性为基础的指标，设计多视角Top-K排序进行多项相关工作的评估。本项目设计多视角热点新闻推送和交互系统。

­­­

**项目论证：**

1. 研究现状和研究意义

新闻推送系统如今被大量使用在各大主流媒体平台，如国内的头条、微博、腾讯新闻等；国外的Google News、Twitter等，然而其通常依赖于相对单一的指标，如用户历史记录等进行高度个性化的推送，导致用户可以接触到的资讯范围被限制在属于用户自己的小圈子里，新闻资讯呈现出“千人千面”的特点。用户在浏览热点新闻时，随着对某一观点的频繁点击，基于单一指标的top-k会趋向于继续推送片面的新闻，这就导致用户对热点新闻的认识偏差。由此可见设计一款基于多视角top-k的新闻推送系统对于全面提高新闻推送质量至关重要。

知识图谱是一种异构图，其中结点表示实体，边表示实体间的关系。近年来，知识图谱作为一种结构化的人类知识形式，引起了学术界和工业界的极大关注。目前知识图谱的应用已经较为成熟。其中KGE算法是知识图谱表示为地位的嵌入向量，分为基于距离和基于语义匹配的模型，基于距离的模型有TransE，TransH，TransR，TransD等；而基于语义匹配的模型有DistMult等。基于知识图谱的算法具有可解释性强的优点，因为它采用实体-关系-实体的结构化方式来表达知识。同时，这种结构让知识图谱易于查询、拓展、可视化展示，十分适合用于多视角新闻推送系统的设计中。

文本情感分析又称意见挖掘, 是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程。目前，情感分析技术可以分为两类，一类是基于机器学习的方法，另一类是基于情感词典的方法。

目前已经有大量公开的情感词典供研究者使用，在基于传统机器学习方法和基于深度学习的方法中有研究者使用公开发表的数据集，也有研究者通过构建新数据集，划分不同的标签来实现对于情感的分类。心理学中定义了七种基本的情感：恐惧、愤怒、喜悦、悲伤、蔑视、厌恶和惊讶，可以将不同新闻文本分成这七种类别。

KGAT（Knowledge Graph Attention Network for Recommendation）是推荐系统中一种基于知识图谱和注意力机制的新方法。现有的CKG模型在利用高阶信息进行推荐时，面临着节点规模过大、高阶关系对预测贡献不平等的挑战。KGAT通过引入线性时间复杂度的递归嵌入传播和基于注意的聚合，利用神经注意机制来学习传播过程中每个邻居的权重，试图解决这些挑战。

1. 难点
   * + 1. 如何对用户的行为多视角分析并多视角挖掘新闻

新闻推送过程中，用户的行为模式（例如：点击、滚动、停留、点赞、转发等）包含了很多用户对于当前新闻推送内容的个性化评价信息，这些信息通常是被系统所忽略的，如何利用这些行为信息分析用户并且构建用户画像从而挖掘多视角新闻内容是一大难点。

* + - 1. 如何对新闻推送各个指标做综合排序

在新闻推送过程中，准确性、时效性和全面性是三个重要的指标，但它们往往会存在冲突。三项指标该以如何的权重占比进行综合排序是一个难点。此外，对于各种不同类型的新闻内容，三项指标的占比是否应该体现出差异性，这个差异性具体如何实现，这也是新闻综合排序的难点之一。

1. 研究内容和创新之处

知识图谱注意力模型

本项目引入KGAT模型，KGAT通过将用户和物品的特征表示为多维向量，并利用知识图谱中的实体和关系来构建用户-物品交互矩阵，然后采用注意力机制来自适应地聚合用户和物品的向量表示，以更好地捕捉它们之间的关系和相互作用。本项目在KGAT模型的基础上，以端对端的方式显式地模拟CKG中的高阶连接，量化分析用户的行为（例如：点击、滚动、停留、点赞、转发等），充分利用知识图谱中的上下文信息和图神经网络中的新闻特征，从而通过用户的行为挖掘更深层次的相关内容。

多视角top-k排序方法

本项目利用知识图谱设计多视角注意力模型，综合新闻文本的准确性、时效性和全面性，综合设计top-k指标，在神经网络中进行大规模数据训练。为进一步加强模型性能，本项目在BERT的基础上结合NLP自然语言处理技术和情感词典，可以做到更加精准地对新闻内容和观点进行分类。

1. 预期成果
2. 提出基于知识图谱的­­­多视角注意力模型

为了能在保证准确度的同时给用户有一种新颖的新闻推送体验，本项目提出一种多视角注意力模型。传统的新闻推送模型只从一个方面考虑新闻内容，而该模型则会同时考虑多个视角，例如新闻的点击率，火爆程度，时效性以及观点全面性等。这种设计可以更全面地了解新闻事件，提高推送的准确性。在考虑多视角时，该模型通过注意力机制来加权不同视角的重要性，从而更好地平衡不同视角对推送的影响。这种设计可以更好地控制时效性和全面性的平衡。该模型不仅考虑新闻事件的时效性和全面性，还结合了事件的准确性，例如通过对不同媒体的报道进行比对，来确定新闻事件的真实性。这种设计可以提高推送的准确性和可信度。

1. 设计综合top-k指标

本项目在实现上述模型过程中设计综合top-k指标，综合考量新闻的时效性，准确性和全面性。本项目引入KGAT框架，KGAT可以通过学习用户和新闻事件的表示向量，使用注意力机制来加权不同实体之间的交互，从而提高推送的准确性和个性化程度。在KGAT模型的基础上，以端对端的方式显式地模拟CKG中的高阶连接，充分利用知识图谱中的上下文信息和图神经网络中的新闻特征，给用户top-k推送，并保证了新闻推送的准确度和时效性。针对新闻全面性，本项目引入BERT模型，把新闻的标题，正文和事件三者量化并融合，得出新闻的全面性指标。为进一步加强模型性能，本项目在BERT的基础上结合NLP自然语言处理技术和情感词典，可以做到更加精准地为各条新闻做情感分类。本项目所设计的top-k指标融合了上述关于新闻准确性，时效性，全面性，从而给推送系统提供一个可靠的参考指标。

1. 设计多视角新闻推送服务

本项目在上述模型及top-k指标的基础上，设计了一种可视化多视角新闻推送服务系统，从热点事件出发，基于上述标准给用户推送新闻。本项目在移动端为用户设计一套可视化交互系统。根据热点事件，系统会推送与之相关的7个关键词，用户此时可以根据自己的兴趣选择关键词，之后系统会按照本项目中设计的综合top-k指标以及多视角注意力模型推送一系列新闻。本系统的优势就在于没有完全按照用户的个人喜好来做新闻推送，用户的兴趣在整个推送过程中被减少，缓解了用户获取资讯范围会逐渐减小的问题。此外，综合top-k的相关数据也会在对应新闻旁展示给用户。得益于知识图谱的应用，用户也可以清晰地了解到自己为什么会收到这些新闻推送。

1. 研究方法

本项目主要采用以下研究方法：1，理论研究，先进行资料查阅和技术学习，通过各种文献以及网上资料，对推荐系统的关键算法进行初步整合与实现，并记录各种当前问题和潜在问题。2，现有模型研究，如GCN、GAT、GAE、GRCN、GTN等，通过对当下现有的新闻推荐系统进行实验研究，参考各大主流的新闻推荐系统，从中得到相关的启发和提升用以完善新闻推荐系统。3，定性定量研究，由于新闻推荐需要考虑的因素有很多，诸多因素错综复杂，需要将各个因素对该系统的影响做出准确判断，以此来决定各种因素在该系统的权重。4，设计多视角注意力模型，将新闻的时效性，准确性和全面性进行建模。

1. 参考文献

[1]刘丹阳. 结合知识图谱的个性化新闻推荐系统[D].中国科学技术大学,2022.DOI:10.27517/d.cnki.gzkju.2022.000773.

[2]王鸿伟. 基于网络特征学习的个性化推荐系统[D].上海交通大学,2018.DOI:10.27307/d.cnki.gsjtu.2018.000175.

[3]唐柳. 基于知识图谱的个性化农业新闻推荐系统研究[D].安徽农业大学,2022.DOI:10.26919/d.cnki.gannu.2022.000562.

[4]秦川,祝恒书,庄福振,郭庆宇,张琦,张乐,王超,陈恩红,熊辉.基于知识图谱的推荐系统研究综述[J].中国科学:信息科学,2020,50(07):937-956.

[5]朱冬亮,文奕,万子琛.基于知识图谱的推荐系统研究综述[J].数据分析与知识发现,2021,5(12):1-13.

[6]田萱,陈杭雪.推荐任务中知识图谱嵌入应用研究综述[J].计算机科学与探索,2022,16(08):1681-1705.

[7]刘春霞,武玲梅,谢小红.推荐系统评估研究综述[J].现代计算机(专业版),2018(24):11-15+20.

[8]周万珍,曹迪,许云峰,刘滨.推荐系统研究综述[J].河北科技大学学报,2020,41(01):76-87.

[9]赵岩,刘宏伟.推荐系统综述[J].智能计算机与应用,2021,11(07):228-233.

[10]常亮,张伟涛,古天龙,孙文平,宾辰忠.知识图谱的推荐系统综述[J].智能系统学报,2019,14(02):207-216.

[11]Song W, Duan Z, Yang Z, et al. Explainable knowledge graph-based recommendation via deep reinforcement learning. 2019. ArXiv: 190609506

[12]Wang X, He X N, Cao Y X, et al. KGAT: Knowledge Graph Attention Network for Recommendation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 950-958.

[13]Wang H W, Zhang F Z, Xie X, et al. DKN: Deep Knowledge Aware Network for News Recommendation[C]//Proceedings of the 2018 World Wide Web Conference. 2018: 1835-1844.

[14]范琳娟,孙喁喁,徐飞,周行行.注意力与多视角融合的新闻推荐算法[J].计算机系统应用,2022,31(12):178-186.DOI:10.15888/j.cnki.csa.008890.

[15]白刚.基于语义与情感词典的微博评论情感分析方法[J].现代计算机,2021,27(30):55-58+63.

**项目方案：**

**训练目标：**

1. 实现一个基于用户个性化的新闻推送系统。此系统在具备一般新闻推荐系统的基础功能的基础上，应能进一步提高用户的新闻阅读效率，解决当下一部分新闻推荐系统存在的新闻时效性不足、全面性不够等问题。在具体实现的层面上，首先需要学习和完善推荐系统的算法，多多了解各种新闻推荐算法的优缺点；然后选择和完善适合的知识图谱，学习运用神经网络的相关知识，将其用合适的算法应用进系统，并最终实现相对完善的推送系统。
2. 项目组的成员在实现项目的过程中，学习推荐系统、卷积神经网络、知识图谱等有关机器学习、深度学习的知识。进一步掌握编程软件的应用。了解计算机理论知识在实际情境中的应用，为日后的学习以及工作或研究打下基础，并提高对理论知识的应用意识。在实践过程中锻炼信息检索能力、学习能力、时间管理能力、语言组织能力以及合作能力等。

**前期准备：**

1. 多角度调研。在学术上，首先需了解该领域的基本常识和已有成果。查阅技术论坛文章、论文等，并结合一些已有的代码。在阅读学习的过程中，明悉、细化并结构化不同的概念，为后续的实践步骤打下坚实的理论基础。在应用上，需要考察、研究实际的推荐场景。由于我们选定新闻为推送系统的实体，就需要了解市面上的新闻网站、新闻软件的推荐逻辑（算法实现方式）以及知识库等。同时，为提升用户体验，可以用网络信息检索等方式了解用户的实际需求，为后续的算法设计做准备。
2. 在调研的同时，完善工具应用方面的知识漏洞。如继续学习python语法，了解深度学习的有关知识，学习tensorflow的使用，熟悉虚拟环境的搭建等。
3. 阅读现有DKN模型代码，结合先前了解的理论知识，尝试读懂代码、理解算法，并在该过程中进一步理解理论概念。
4. 在理论理解和代码阅读的基础上，逐步确定具体化的、当前可实现的目标。分清不同目标的权重，重点目标应当首要实现，次要目标则根据能力和系统限制，尽量实现。在此认知基础上实现和完善项目代码。

**组织实施：**

1）项目组的各成员都应提高个体参与度。细分任务，合理分工，明确小任务的截止日期。定期用线上线下沟通渠道讨论、汇报当下进度，并根据具体情况灵活调整分工和阶段性任务，以提高项目组织的效率。

2）平时加强沟通交流，发挥团队作用，遇到疑难进行讨论互助。

3）向导师定期进行项目进度汇报。必要时可寻求方向性或者细节指导。

**安成项目的条件和保障：**

1.本项目申报团队中的成员积极性很高，学习能力很强，能够充分安排和利用课余时间来完成对本项目的研究和创新。指导老师和研究生学长专业知识储备丰富，和团队成员之间交流顺畅，为研究指明方向。

2.学校提供高性能运算服务器，保证模型训练过程顺利。

**经费预算：**

1.论文发表版面费：用于发表论文或申请软件专利，预计4000元

2.差旅费：在项目实施过程中，开展业务调研等，预计1000元

3.合作与交流费:用于与资助项目研究工作直接相关的合作与交流费用，包括项目组人员的参观、培训及相关专家的访问等部分费用，预计2000元

4.打印费：预计500元

5.资料费：用于购买项目模型训练中使用到的数据，预计1000元

6.能源成本:项目过程中，相关大型仪器设备和专用科学装置运行过程中产生的水、电等费用，预计500元

总计：9000元

。