

29 | 推荐系统的工程化策略及服务部署策略

2023-06-21 黄鸿波 来自北京

《手把手带你搭建推荐系统》

shikey.com转载分享

你好，我是黄鸿波。

到现在，我们基本上已经完成了推荐系统中大部分的内容，这节课我们来聊一聊推荐系统在企业中上线的一些策略，以及与服务部署的相关方案。

我将本节课分为了下面两大模块。

1. 推荐系统中推荐列表组成策略。
2. 推荐系统的服务部署方案介绍。

下面正式开始今天的内容。

推荐列表组成策略

首先来看看推荐系统中的推荐列表到底是怎么来的。

当我们使用各种 App 或者网站时，为了使用户的黏性更高，推荐的内容似乎就是由推荐系统中的召回集产生然后再经过排序层进行排序所产生的，这样可以确保用户大概率会看到喜欢的内容。

shikey.com转载分享

但是这样会产生一个问题：推荐到最后内容会越来越聚焦，内容类别越来越少（[🔗 EE 问题](#)）。

解决这个问题的方法有很多，比如增加多样性、使用一些 Bandit 算法等。最简单的方法是在推荐系统最后输出的排序列表中加入一些新的内容，丰富推荐列表。

还有另外一个问题，App 的第一个位置往往并不是我们喜欢的内容，对于新闻类的 App 大部分的情况下是当下的时事热点放在前面；对于产品类的 App（比如说某款游戏的相关工具 App），那么前面很有可能是公告或者活动相关的信息；对于电商类 App，前面往往可能放的是一些大品牌的商品，但是这个类别也许你从来都没有接触过。在推荐系统中这一类的推荐叫做运营推荐，运营推荐的位置叫做运营推荐位。

到这里你能看出，现在的推荐列表由下面这个公式组成。

推荐列表 = 运营推荐位 + 排序算法位 + 最新内容位（也可以称为最新池）

借着这个公式，我们再来对排序算法位做进一步的拆解。

排序算法实际上是从召回集中获取内容，也就是说，召回集的质量往往也决定着排序层输出的内容的质量，也就决定着最后的点击率等相关问题。那么召回集是怎么来的呢？

在真实的企业案例中，召回集往往是由多个召回算法以一定的形式组合在一起，这里实际上又会存在一个召回算法的选择问题。

假设现在的召回集由协同过滤、LFM、YouTubeDNN、基于时间的召回、基于热度的召回这五种召回算法组成，我们可以把这几种召回算法的结果融合在一起，形成一个新的召回集（因

为这些召回算法在不同的场景下具有不同的优缺点)。具体来说,分为下面三步。

1. 为每个召回算法赋予一个权重, 通常使用经验权重或者机器学习等方法来确定。
2. 利用各个召回算法的结果和对应的权重, 构建一个加权召回结果。
3. 对加权召回结果进行排序, 取 Top N 个物品作为最终的召回结果。

在构建召回集的过程中, 你需要注意各个召回算法的权重应该根据实际情况而定, 可以通过日志数据、A/B 测试等手段来识别最优权重。此外, 不同的召回算法可能存在重复召回或遗漏召回的情况, 需要多多关注。

除了加权融合策略, 还有一些其他的召回集构建方法(如基于组合优化、集成学习等), 这些方法的选择应该根据实际场景和需求而定。同时, 还需要注意动态调整召回算法和召回集的构建方式, 以保持推荐系统的稳定性和准确性。

上面说的是多样性召回怎么做, 实际上除了多样性召回之外, 我们也可以采用多样性排序。一般多样性排序可以用下面五种方式。

1. **线性组合**: 将多个排序算法的排序得分按照权重加权相加, 得到新的排序得分。权重可以通过统计实验或者人工指定, 以适应不同场景下的需求。
2. **基于遗传算法的融合**: 使用遗传算法求解多个排序算法的权重, 遗传算法使用交叉、变异和选择等过程, 将每个排序算法的权重当作染色体的基因, 根据适应度函数来更新染色体的值, 最终得到权重最优化的结果。
3. **级联模型**: 将多个排序算法串联起来, 将前一次排序得到的结果进一步转换为下一级排序算法的输入, 直到得到最终的排序结果。
4. **非线性叠加模型**: 通过一些函数或者神经网络模型将多个排序算法融合在一起, 得到新的排序得分, 然后根据该得分进行排序。
5. **排序组合**: 将多个排序算法根据不同的情况和需求进行组合, 例如根据不同的时间段使用不同的排序算法, 或者根据不同的用户行为使用不同的排序算法等等。

需要注意，选择哪种方法取决于应用场景和数据特征。同时，需要在实验中评估不同的排序方法，选择最适合的方法进行使用。

最终可以得出公式如下。

推荐列表 = 运营推荐位 + 多样性排序算法位（包括多样性召回） + 最新内容位（最新池）

这样就可以保证在推荐系统中既能推荐出比较新的内容，又能符合用户的口味，是推荐系统工程中常见的策略。

推荐系统的服务部署方案

讲完推荐列表组成策略之后，接下来我们来看看推荐系统的服务部署方案。

从宏观上来讲，推荐系统的服务部署方案有下面五种。

1. 云部署

云部署是将推荐系统服务部署在云端基础设施上，如阿里云、亚马逊 AWS、微软 Azure 等。通过云部署可以快速搭建推荐系统，并且可以根据业务需求进行扩展和收缩，还可以提高系统的可靠性和灵活性。

2. 本地部署

本地部署是指将推荐系统服务部署在本地服务器或者集群上，一般需要进行硬件、软件等基础设施的采购和配置。本地部署可以根据业务需求进行定制，同时也可以提高系统的可控性和灵活性。

3. 容器化部署

容器化部署是将推荐系统服务及其依赖项打包成容器，使用 Docker 等容器平台进行部署和管理。容器化部署可以简化系统的安装和配置，提高系统的可移植性和部署效率。

4. Serverless 部署

Serverless 部署是一种无服务器的部署方式，通过使用云提供商的函数计算等服务来部署推荐系统，避免了硬件和软件的采购和配置，更加简单和轻量级，但需要根据业务场景和功能需求来选择服务商和功能模块。

shikey.com转载分享

5. 分布式部署

分布式部署是将推荐系统服务进行分布式部署，通过将不同的服务分配到不同的节点上来提高系统的吞吐量和可靠性。分布式部署需要采用一些分布式技术如 Zookeeper、Dubbo 等进行服务发现和调度。

一般来讲，如果是一个小型团队或者初创公司，由于没有过多的资金和精力去采购和维护一台服务器，所以一般采用云部署。但云部署和容器化部署可以结合使用，我们可以用下面的六步将它们结合。

1. **确定容器平台和云服务商**：首先需要选择一个容器平台和一个云服务商，并将相应的账号和密钥进行配置和设置。
2. **构建容器镜像**：将推荐系统的代码库打包成镜像，可以通过 Dockerfile 构建，也可以使用开源工具或专业工具进行打包，将打包好的镜像上传到 Docker 仓库。
3. **配置容器运行环境**：根据推荐系统的需要进行对应的环境配置，例如配置数据库、缓存、消息队列等，可以使用云服务商提供的相关服务进行配置。
4. **部署容器**：使用云服务商提供的容器编排工具，如 Kubernetes、Docker Swarm 等，将容器镜像部署到云端服务器上，并进行相应的扩容和缩容等操作。
5. **监控和日志管理**：使用云服务商提供的监控和日志管理工具，对推荐系统进行监控和管理保证应用的正常运行，及时处理异常情况。
6. **系统维护和更新**：定期维护和更新推荐系统，并对容器和云服务进行升级和优化，提高系统的性能和稳定性。

总之，推荐系统中的云部署和容器化部署结合，可以实现自动化部署、高可用性、弹性伸缩和资源利用率最大化，是一个高效、灵活、可靠的应用部署和管理方式。

如果是在一个有本地机房的团队中，更加推荐使用本地化部署加容器化部署的方案。原理和云部署加容器化部署一样，但本地化部署是将推荐系统程序部署在本地服务器或本地网络中，相比于云部署来说，有下面五点优势。

1. **安全性更高**：本地化部署有更好的数据安全性，因为数据不会被传输到公共云服务商，数据也更容易被保护和管理。shikekey.com转载分享
2. **相对稳定的网络环境**：本地化部署通常能够提供更稳定的网络环境，因为其不受云服务商的网络延迟、带宽限制、网络中断等影响。
3. **更多的自主权**：本地化部署可以更好地适应企业自身的需求，有更大的自主权，企业可以根据自身需求自定义硬件配置，选择合适的软件架构等。
4. **更好管理控制**：本地化部署可以更方便地管理和控制整个系统，包括监控系统运行状态、修改应用配置、优化系统性能等。
5. **操作成本更低**：本地化部署不存在云服务商的服务器租赁费用和带宽流量费用，因此成本更低，系统的运维也更为灵活和自由。

总之，本地化部署与云部署相比具有更高的安全性、更稳定的网络环境、更多的自主权、更好管理和控制，以及更低的操作成本。但本地化部署也面临着硬件维护成本高、系统可扩展性差、更难应对突发流量等挑战，企业应根据自身的需求和实际情况选择合适的部署方式。

总结

本节课主要是带你学习了一些理论性的内容，讲解了推荐系统的推荐列表策略和部署方案，下面我们来做一个简单的总结，学完本节课你应该知道以下四个要点。

1. 一个企业级的推荐系统列表是由运营推荐位 + 多样性排序算法位 + 最新内容位组成。
2. 在进行推荐系统的列表排序前应该使用多种策略进行融合，这样可以使推荐的内容既丰富又满足用户的需求。
3. 推荐系统的部署方案有云部署、本地化部署、容器化部署、Serverless 部署、分布式部署等等。
4. 一般情况下我们会采用多种部署方案结合的方式进行推荐系统部署。

课后思考

学完本节课，给你留一个思考题：你会用哪种方式部署产品，为什么？

期待你的分享，如果今天的内容让你有所收获，也欢迎你推荐给有需要的朋友！

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

shikey.com 转载分享

精选留言

由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。