

# 基于KL散度与JS散 度相似度融合推荐算

汇报人：

法

2024-01-24



# 目录

- 引言
- KL散度与JS散度理论基础
- 基于KL散度和JS散度的相似度度量方法
- 融合推荐算法设计
- 实验结果与分析
- 总结与展望



01

引言





## 背景与意义



互联网信息爆炸式增长，用户面临信息过载问题，个性化推荐系统应运而生。



推荐系统核心在于准确度量用户与物品之间的相似度，KL散度与JS散度作为相似度度量方法，在推荐算法中具有广泛应用。



基于KL散度与JS散度相似度融合推荐算法能够综合利用两种散度的优势，提高推荐准确性，优化用户体验，具有重要的现实意义。



# 国内外研究现状



国内外学者在推荐算法领域进行了大量研究，提出了基于内容、协同过滤、深度学习等多种推荐算法。

KL散度与JS散度在推荐算法中的应用得到了广泛关注，但单一使用某种散度可能存在局限性。



目前已有一些研究尝试将不同相似度度量方法进行融合，以提高推荐性能，但仍存在融合方式单一、效果不稳定等问题。



# 本文研究内容与创新点

01

研究内容：本文提出了一种基于KL散度与JS散度相似度融合的推荐算法。首先，利用KL散度度量用户与物品之间的相似度；其次，运用JS散度对KL散度结果进行修正；最后，通过融合两种散度的相似度结果生成推荐列表。

02

创新点

03

1. 提出了一种新的相似度融合方式，综合利用了KL散度与JS散度的优势。

04

2. 通过实验验证了所提算法在推荐准确性、覆盖率和多样性等方面的性能表现。

05

3. 探讨了不同参数设置对算法性能的影响，为实际应用提供了参考依据。

02

# KL散度与JS散度理论基 础





# KL散度定义及性质



01

定义：KL散度  
( Kullback-Leibler Divergence ) 是一种衡量两个概率分布P和Q相似度的非对称度量方法，记作 $D_{KL}(P||Q)$ 。其计算公式为： $D_{KL}(P||Q) = \sum [P(x) * \log(P(x) / Q(x))]$ ，其中x为所有可能的事件。

02

性质

03

1. 非负性： $D_{KL}(P||Q) \geq 0$ ，当且仅当 $P=Q$ 时取等号。

04

2. 不对称性： $D_{KL}(P||Q) \neq D_{KL}(Q||P)$ ，即KL散度不满足距离度量的对称性。

05

3. 凸性：对于固定的Q， $D_{KL}(P||Q)$ 是关于P的凸函数。







# JS散度定义及性质



01

定义：JS散度 ( Jensen-Shannon Divergence ) 是一种基于KL散度的对称相似度量方法，记作 $D_{JS}(P, Q)$ 。其计算公式为： $D_{JS}(P, Q) = 1/2 * D_{KL}(P||M) + 1/2 * D_{KL}(Q||M)$ ，其中 $M = 1/2 * (P + Q)$ 为P和Q的均值分布。

02

性质

03

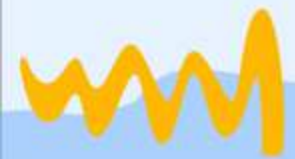
1. 非负性： $D_{JS}(P, Q) \geq 0$ ，当且仅当 $P=Q$ 时取等号。

04

2. 对称性： $D_{JS}(P, Q) = D_{JS}(Q, P)$ ，满足距离度量的对称性。

05

3. 平方根性质： $D_{JS}(P, Q)$ 的平方根满足距离度量的三角不等式。





# KL散度与JS散度关系分析



## 联系

KL散度和JS散度都是用于度量两个概率分布相似度的方法，且都以信息熵为基础。其中，JS散度是在KL散度的基础上进行了对称化处理。

## 1. 对称性

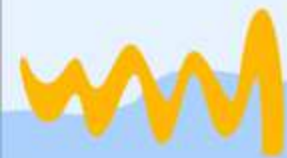
KL散度是非对称的，而JS散度是对称的。这使得JS散度在度量两个分布相似度时更加公平。

## 2. 取值范围

KL散度的取值范围为 $[0, +\infty)$ ，而JS散度的取值范围为 $[0, 1]$ 。这使得JS散度的数值更易于理解和比较。

## 3. 应用场景

由于KL散度和JS散度的特性不同，它们在实际应用中的选择也有所差异。例如，在机器学习和信息论中，KL散度常用于度量模型预测分布与真实分布之间的差异；而JS散度则常用于文本相似度匹配、图像分割等任务中。



03

**基于KL散度和JS散度的  
相似度度量方法**





# 数据预处理及特征提取方法



## ● 数据清洗

去除重复、无效和异常数据，保证数据质量。

## ● 特征提取

从原始数据中提取出能够反映用户兴趣和物品属性的特征，如用户历史行为、物品标签等。

## ● 数据标准化

消除特征间的量纲差异，使得不同特征具有可比性。





# 基于KL散度的相似度度量方法



## KL散度定义

KL散度是衡量两个概率分布P和Q差异的一种方法， $D(P||Q)$ 表示用概率分布Q来近似概率分布P时所造成的信息损失。

## 相似度计算

将用户-物品交互数据转化为概率分布，利用KL散度计算用户或物品间的相似度。值越小，相似度越高。

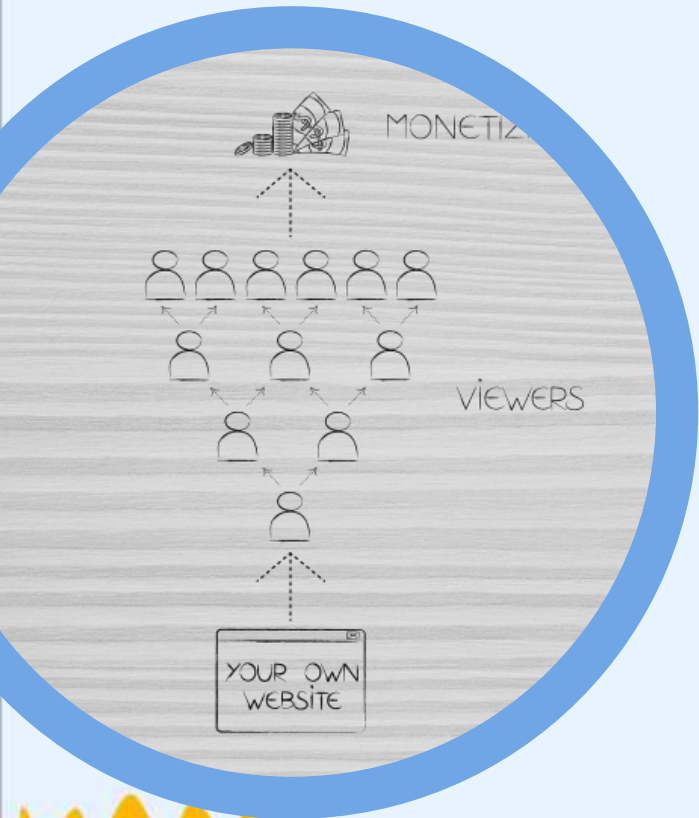
## 优缺点分析

KL散度能够衡量分布间的差异，但不对称且不满足三角不等式，使用时需注意。





# 基于JS散度的相似度量方法



## JS散度定义

JS散度是衡量两个概率分布相似度的一种方法，基于两个分布的平均值进行度量，具有对称性。

## 相似度计算

将用户-物品交互数据转化为概率分布，利用JS散度计算用户或物品间的相似度。值越小，相似度越高。

## 优缺点分析

JS散度具有对称性且取值范围在 $[0,1]$ 之间，便于比较和解释。但在某些情况下，如两个分布完全不重叠时，JS散度可能无法提供有效的相似度量。

04

# 融合推荐算法设计





# 融合策略选择与优化



01

策略一：线性加权融合

02

对KL散度和JS散度计算出的相似度进行线性加权，得到综合相似度。

03

权重可根据实际数据分布或经验进行调整。

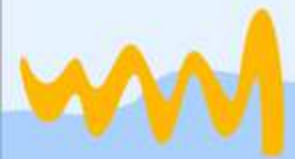
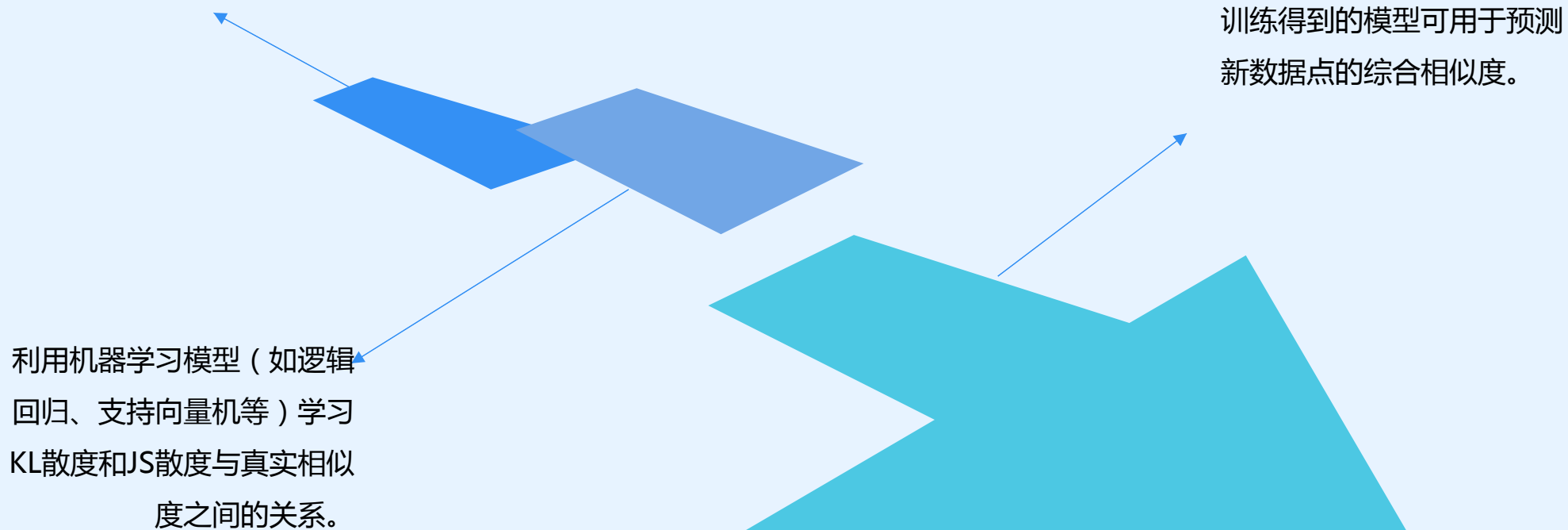




# 融合策略选择与优化



策略二：基于机器学习的融合





# 融合策略选择与优化



01

优化方法：交叉验证与网格搜索

02

通过交叉验证评估不同融合策略的性能。

03

利用网格搜索确定最佳参数组合，如线性加权的权重或机器学习模型的超参数。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：  
<https://d.book118.com/135304311241011230>