基于最大偏差相似性准则的交通流聚类算法*

黄何列，蔡延光，蔡 颢，戚远航
（广东工业大学 自动化学院，广州 510006）

摘 要：针对常用聚类算法对随机性强、波动频繁的交通流聚类效果不理想的问题，提出了一种新的交通流相似性度量准则——最大偏差相似性准则，并提出了一种基于最大偏差相似性准则的交通流聚类算法。最大偏差相似性准则能够有效刻画频繁波动交通流曲线的形态相似性，具有简明、合理、灵活等特点；聚类算法无须预先指定类别数，能够保证类间曲线的明显差异性和类内曲线的高度相似性。实验表明，所提出的算法聚类效果明显优于常用聚类算法，聚类结果能够较好地满足实际应用的需要。

关键词：交通流曲线；聚类算法；曲线形态；相似性
中图分类号：TP274 **文献标志码**：A **文章编号**：1001-3695(2018)08-2274-03
doi:10.3969/j.issn.1001-3695.2018.08.008

Traffic flow clustering algorithm based on maximum deviation similarity criterion

Huang Helie, Cai Yanguang, Cai Hao, Qi Yuanhang
(*School of Automation*, *Guangdong University of Technology*, *Guangzhou* 510006, *China*)

Abstract: Focusing on the problem that the common clustering algorithms are not ideal for traffic flow with strong randomicity and frequent fluctuation, this paper proposed a new traffic flow similarity measurement which was called maximum deviation similarity criterion (MDSC), and proposed a traffic flow clustering algorithm based on the MDSC. The MDSC could effectively describe the curve shape similarity of frequent fluctuating traffic flow, which had the characteristics of simple, reasonable, flexible and so on. The proposed clustering algorithm did not need to specify the number of classes in advance, which could ensure that the curves of different classes have obvious differences and the curves in the same class have high similarity. The experiments show that the clustering effect of the proposed algorithm is significantly better than that of the common clustering algorithms, and the clustering result of the proposed algorithm can better meet the needs of practical applications.

Key words: traffic flow curve; clustering algorithm; curve shape; similarity

随着智能交通系统的高速发展，积累了大量的交通流检测数据，其中蕴涵着大量有价值的知识[1]。交通流聚类就是在大量的交通流检测数据之上，利用适合的聚类算法将交通流曲线形态相似的日期或路段归为一类，其目的是帮助交管部门、交通信息服务提供商准确掌握交通流的变化趋势和典型模式，为更好地实现交通预测、交通控制和交通管理等提供重要支持[2~6]。在智能交通系统中，交通流是最常见的时间序列数据。对时间序列数据聚类的常用聚类算法有 K-均值聚类算法[7]、K-中心点聚类算法[8]、模糊 C-均值聚类算法(fuzzy C-means，FCM)[9]、层次聚类算法[10]、基于密度的聚类算法[11]，以及一些在以上算法基础上改进的算法[12~14]。Huang 等人[15]首先采用 FCM 聚类算法对交通流进行聚类，获取交通流的典型模式，然后在聚类结果的基础上进行交通流预测，提高了预测的精度和鲁棒性。李清泉等人[16]通过 K-均值聚类算法获得交通流的典型模式并对其时变特征进行了分析。闫伟等人[17]探讨了 FCM 聚类算法在交通流聚类中的应用，利用蚁群算法初始化模糊隶属度函数和聚类中心，提高了 FCM 算法对交通流的聚类性能。Jiang 等人[18]首先用 K-均值算法将不同节点的交通流时间序列聚类，然后分析了不同节点交通流的波动模式以及关联性。上述聚类算法通常采用欧氏距离作为相似性度量准则，但由于交通流时间序列数据通常具有高维、高噪声、波动频繁等特点，这样有可能把一些在形态上差异明显的交通流曲线聚到同一类中，使得聚类结果实用性较差；其次，很多聚类算法要求预先设定类别数，而类别数的不正确设定可能导致聚类的不合理性；此外，上述聚类算法虽然满足了某些实际应用的需要，但缺乏一定的灵活性，聚类精度没有多少选择余地，不能满足有不同相似性精度要求的实际应用需要。

为了更好地对随机性强、波动频繁的交通流进行聚类，本文提出了一种基于最大偏差、相似度和偏离度的交通流相似性度量准则——最大偏差相似性准则，并提出了基于最大偏差相似性准则的交通流聚类算法。所提出的最大偏差相似性准则能够抑制交通流频繁波动对相似性度量的不利影响，有效地刻画同类交通流曲线的形态相似性；尤其是最大偏差、相似度和偏离度的选择自由度保证了最大偏差相似性准则具有较大的灵活性，能够对曲线的相似性进行多精度度量。实验表明，所提出的算法聚类效果明显优于 K-均值聚类算法、K-中心点聚类算法和 FCM 聚类算法，具有较好的实用性。

1 相似性度量准则

1.1 交通流数据归一化

本文将按照交通流时间序列数据的曲线形态进行聚类，所以需要预先对其进行归一化处理。设有 n 个 m 维的交通流时间序列数据，第 i 个交通流时间序列数据为 $X_i=(X_{i1},X_{i2},\cdots,X_{im})$，$X_{ik}(\geq 0)$ 为第 i 交通流时间序列数据第 k 个时间点的交通流量，$i=1,2,\cdots,n,k=1,2,\cdots,m$。其中，$m$ 的值不能过小，至少在 10 以上。交通流时间序列数据的归一化处理公式如下：

$$x_{ik}=\frac{X_{ik}}{\max\limits_{1\leqslant k\leqslant m}\{X_{ik}\}} \tag{1}$$

显然，x_{ik} 的值域为 $0\leqslant|x_{ik}|\leqslant 1,i=1,2,\cdots,n,k=1,2,\cdots,$

收稿日期：2017-03-28；**修回日期**：2017-05-04 **基金项目**：国家自然科学基金资助项目(61074147)；广东省自然科学基金资助项目(S2011010005059)；广东省教育部产学研结合项目(2012B091000171,2011B090400460)；广东省科技计划项目(2012B050600028,2014B010118004,2016A050502060)；广州市花都区科技计划项目(HD14ZD001)；广州市科技计划项目(201604016055)

作者简介：黄何列(1990-)，男，四川射洪人，硕士，主要研究方向为数据挖掘、智能交通(m15915854571@163.com)；蔡延光(1963-)，男，教授，博士，主要研究方向为人工智能、智能决策系统；蔡颢(1987-)，男，研究员，博士，主要研究方向为大数据分析、智能信息处理；戚远航(1993-)，男，博士，主要研究方向为物流运输调度、智能算法.

m;$x_i=(x_{i1},x_{i2},\cdots,x_{im})$为 X_i 的归一化数据,称为交通流数据,也称为交通流曲线。

1.2　最大偏差相似性准则

为了更好地度量随机性强、波动频繁的交通流曲线之间的相似性,本文提出了最大偏差相似性准则,其具体内容如下:

a)设 s_{ijk} 为 x_i 与 x_j 对应时间点的绝对差值,$i,j=1,2,\cdots,n,k=1,2,\cdots,m$,其计算公式如下:

$$s_{ijk}=|x_{ik}-x_{jk}| \tag{2}$$

b)设 s_{ijk} 满足 $s_{ijk}\leqslant\gamma$ 的个数为 n_{ij},称 n_{ij} 为 x_i 与 x_j 的相似时点数;设 s_{ijk} 连续满足 $s_{ijk}>\gamma$ 的最大个数为 m_{ij},称 m_{ij} 是 x_i 与 x_j 的最大连续偏离时点数,$i,j=1,2,\cdots,n,k=1,2,\cdots,m$。其中,$\gamma$ $(0\leqslant\gamma\leqslant1)$为预设常数,称为最大偏差,它是衡量两个对应时间点交通流相似性的阈值,当 $s_{ijk}\leqslant\gamma$ 时,则认为交通流 x_{ik} 和 x_{jk} 相似,否则不相似;m_{ij} 的具体表达式为

$$m_{ij}=\max\{s\mid\exists k_0,1\leqslant k_0\leqslant m,s_{ijk_0}>\gamma,s_{ij(k_0+1)}>\gamma,\cdots,s_{ij(k_0+s-1)}>\gamma\} \tag{3}$$

c)以交通流数据 x_i 为对比中心,计算 x_j 与 x_i 之间的 n_{ij} 和 m_{ij} $(i,j=1,2,\cdots,n)$,如果 n_{ij} 和 m_{ij} 同时满足下列两个条件:(a)$n_{ij}\geqslant n_0$,其中 $n_0=[\alpha\times m]$,α $(0\leqslant\alpha\leqslant1-m^{-1})$为预设常数,称为相似度;(b)$1\leqslant m_{ij}\leqslant m_0$,其中 $m_0=[\beta\times m]$,β $(m^{-1}\leqslant\beta\leqslant1-\alpha)$为预设常数,称为偏离度。则称交通流数据 x_j 与 x_i 相似,也称为(α,β,γ)-相似,并称条件 a)b)为最大偏差相似性准则。其中,相似度 α 是影响相似曲线形态相似性的重要参数,其值不仅要保证相似曲线形态的高度相似性,又要容许相似曲线之间有一定的较大偏差;偏离度 β 代表了最大偏差相似性准则对相似曲线之间存在连续时间较大偏差的容忍度,其值能够辅助相似度 α 有效刻画频繁波动交通流曲线的相似性。

参数 α、β 和 γ 是决定曲线形态相似性的主要控制参数,当 γ 越小、α 越大、β 越小时,曲线形态相似性越高,反之,则曲线形态相似性越低。但在设置 α、β 和 γ 的值时,并不是 γ 越小、α 越大、β 越小越好,因为那样会使交通流曲线难以找到形态相似的曲线。因此,需要给予它们一定的取值范围,以便它们能适应不同应用场景对曲线形态相似性精度的不同要求。为了使随机性强、波动频繁的交通流曲线能够找到相似曲线,本文设定 α、β 和 γ 的适宜取值分别为:$0.05\leqslant\gamma\leqslant0.25$、$0.7\leqslant\alpha\leqslant0.9$、$m^{-1}\leqslant\beta\leqslant1-\alpha$,通过对它们进行不同的组合可以得到多种较好的相似性精度。

2　算法设计

本文基于最大偏差相似性准则设计了交通流聚类算法,其具体步骤如下:

a)输入 n 个 m 维交通流时间序列数据 $X_i=(X_{i1},X_{i2},\cdots,X_{im})$,$i=1,2,\cdots,n$;输入最大偏差相似性准则参数 α、β、γ;初始化变量、令聚类结果集合 $R=\varnothing$,未聚类的交通流曲线集合 $C=\varnothing$,未聚类的交通流曲线条数 $NC=n$;以 x_i 为对比中心的交通流曲线集合 $S(x_i)=\varnothing$,$S(x_i)$ 中曲线条数 $N(x_i)=|S(x_i)|$,$S(x_i)$ 中所有曲线与 x_i 相似时点总数 $P(x_i)=0$,$S(x_i)$ 中曲线下标集合 $U(x_i)=\varnothing$,$i=1,2,\cdots,n$。

b)按式(1)对 n 个交通流时间序列数据进行归一化处理,得到 n 条交通流曲线 $x_i(i=1,2,\cdots,n)$,并令 $C=\{x_1,x_2,\cdots,x_n\}$。

c)按最大偏差相似性准则计算 n_0、m_0,并对于一切 $i,j=1,2,\cdots,n,j\leqslant i$,计算 n_{ij} 和 m_{ij}。

d)对于一切 $i,j=1,2,\cdots,n$,以 x_i 为对比中心,把 n_{ij}、m_{ij} 分别与 n_0、m_0 比较,将满足最大偏差相似性准则的 x_j 分到 $S(x_i)$ 中,并令 $N(x_i)=|S(x_i)|$、$U(x_i)=U(x_i)\cup\{j\}$、$P(x_i)=\sum n_{ik}$,其中,$k\in U(x_i)$。

e)找出 C 中使 $N(x_i)$ 最大的元素,设 $y_1,y_2,\cdots,y_t$ 为满足这个要求的全部元素;求 x_z,使得 $P(x_z)=\max\{P(y_i)\}$,$i=1,2,\cdots,t,z\in\{1,2,\cdots,n\}$,若有多个满足条件的 x_z,则任选其中一个。

f)如果 $C-S(x_z)=\varnothing$,转步骤 g);否则,对于一切 $x_a\in(C-S(x_z))$,当 $S(x_z)\cap S(x_a)\neq\varnothing$ 时,令 $x_b\in(S(x_z)\cap S(x_a))$,$a,b\in\{1,2,\cdots,n\}$,比较 n_{zb} 和 n_{ab} 的大小。如果 $n_{zb}>n_{ab}$,令 $S(x_a)=S(x_a)-\{x_b\}$、$N(x_a)=|S(x_a)|$、$P(x_a)=P(x_a)-n_{ab}$,否则,令 $S(x_z)=S(x_z)-\{x_b\}$、$N(x_z)=|S(x_z)|$。

g)$R=R\cup\{S(x_z)\}$,$NC=NC-N(x_z)$。

h)如果 $NC=0$,转步骤 i);否则,令 $C=C-S(x_z)$,转步骤 e)。

i)输出聚类结果 R,聚类结束。

3　实验与分析

3.1　实验数据及环境

本文采用实际数据对所提算法进行实验验证。实验数据来源于广深高速某路口 2016 年 5 月共 31 天的入口交通流数据。该路口每天的数据采集从 0:00 开始,到 23:45 结束,采集时间间隔为 15 min,每天采集 96 个交通流数据(单位:辆)。

本文所有实验均在同一实验环境中进行,其中 CPU 为 Intel® Core™ i5-4210M @2.60 GHz,内存为 12 GB,操作系统为 64 位 Windows 7,编程语言为 C++。

3.2　实验结果与分析

本文提出的聚类算法一共有五个参数,分别为 n、m、α、β 和 γ。其中,n、m 的值可直接由实验数据得到 $n=31$,$m=96$;α、β 和 γ 的值是决定算法聚类效果的主要参数,它们的不同组合可以得到不同效果的聚类结果。通过多次实验发现,当 $\alpha=0.8$、$\beta=0.1$、$\gamma=0.15$ 时,本文算法可以取得较好的聚类结果,如图 1 所示。

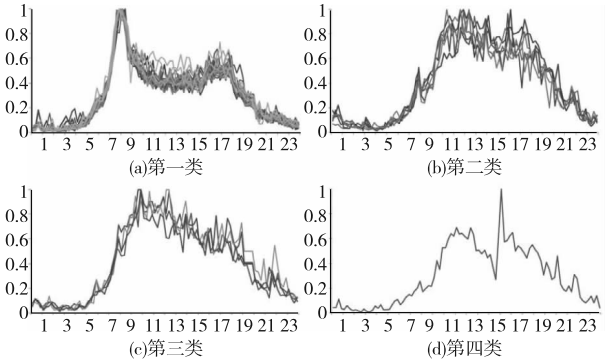

图1　本文算法聚类结果

从图 1 可以看出,在没有设置类别数的情况下,本文算法最终将交通流数据分成了四类,虽然前三类中的个别曲线与其他曲线在一些时段存在一些较大偏离点,但从整体看,同类交通流的曲线形态都非常相似;还可以看出,第四类的交通流属于异常情况,明显不同于其他类别,可以推断该日期属于特殊日期或者出了交通事故等情况。通过查询时间发现,第一类日期都属于工作日,第二类日期都属于星期日,第三类、第四类日期都属于星期六。通过观察各类的曲线形态发现,工作日的交通流变化稳定,有早晚高峰出现,并且早高峰出现在早上 8 点左右,晚高峰出现在下午 5 点左右,与上下班时间较为吻合;还可以发现,节假日的交通流与工作日的交通流区别较大,早高峰明显晚于工作日,这符合人们假日的出行规律;通过对比第二、三类曲线发现,它们在上午的曲线形态非常相似,但是第二类曲线在下午出现了一个小高峰,可能是假期即将结束造成的结果,符合实际情况。通过实验分析可以说明,交通流曲线形态与日期类型关系很大;本文算法将交通流数据聚成四类符合实际情况,聚类效果较好,聚类结果有助于交管部门、交通信息服务商准确掌握各种日期的交通流变化规律。

3.3　对比实验与分析

为了进一步验证本文算法的聚类效果,本节将采用常用的基于欧氏距离的 K-均值聚类算法、K-中心点聚类算法和 FCM 聚类算法对交通流数据进行聚类,并将聚类结果与本文算法比较。

3.3.1　实验结果

K-均值聚类算法、K-中心点聚类算法和 FCM 聚类算法都需要预先设置类别数。通过 3.2 节实验发现,对本文实验数据进行聚类时,类别数设置成四类比较符合实际情况,同时也便

于比较分析。三种对比算法对交通流数据的聚类结果分别如图 2 ~4 所示。

3.3.2 效果分析

1)定性分析

从图 2 ~4 可以看出,三种对比算法聚类结果的第一、二、三类的曲线形态都高度相似,但第四类的个别曲线与其他同类曲线的形态存在明显差异;还可以看出,K-均值聚类算法的前三类、FCM 聚类算法前两类的曲线形态差异都不够明显,应该聚成一类。因此,可以说明三种对比算法虽然具有一定的聚类效果,但是它们都无法保证所有类间曲线具有明显差异性和类内曲线的高度相似性,其聚类效果明显弱于本文聚类算法。

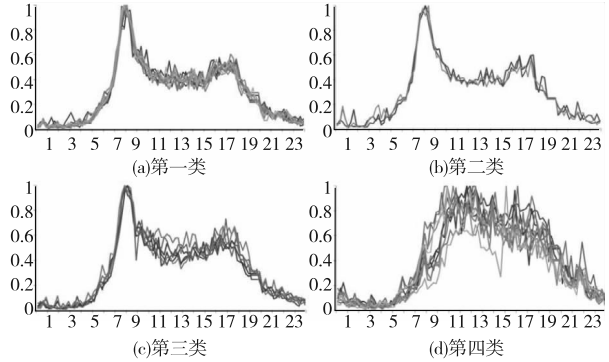

图2 K-均值聚类算法聚类结果

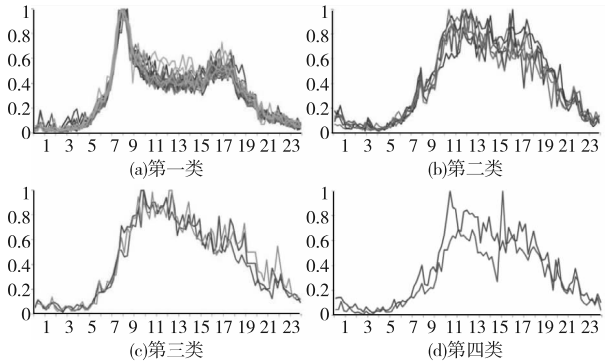

图3 K-中心点聚类算法聚类结果

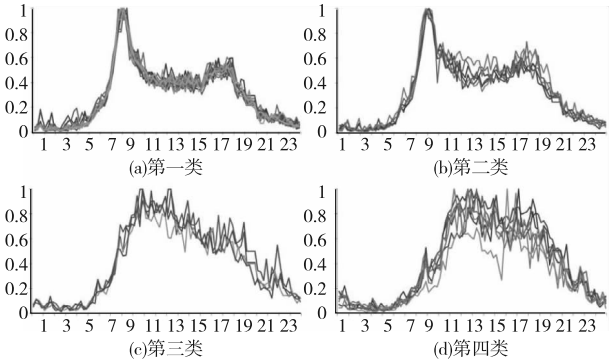

图4 FCM聚类算法聚类结果

2)定量分析 聚类效果定量分析,一般通过考察类的分离情况和类的紧凑情况来评估聚类效果。本文使用所有交通流数据的轮廓系数的平均值来评价聚类效果。

轮廓系数的定义如下:对于有 n 个交通流的数据集 D,假设被划分成 k 个类 $C_1, C_2, \cdots, C_K$。对于每个对象 $x \in D$,计算 x 与 x 所属类的其他对象之间的平均欧氏距离 $a(x)$。类似地,$b(x)$ 是 x 到不属于 x 的所有类的最小平均欧氏距离。设 $x \in C_i (1 \leqslant i \leqslant k)$,则

$$a(x) = \frac{\sum_{x' \in C_i, x \neq x'} \mathrm{dist}(x, x')}{|C_i| - 1} \tag{4}$$

而

$$b(x) = \min_{C_j: 1 \leqslant j \leqslant k, j \neq i} \left\{ \frac{\sum_{x' \in C_j} \mathrm{dist}(x, x')}{|C_j|} \right\} \tag{5}$$

那么,对象 x 的轮廓系数定义为

$$s(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}} \tag{6}$$

$a(x)$ 值的大小反映了 x 所属类的紧凑情况,该值越小,类越紧凑。$b(x)$ 值的大小反映了 x 与其他类的分离程度,该值越大,x 与其他类越分离。$s(x)$ 的值在 $-1 \sim 1$,反映了聚类效果的好坏,该值越大,聚类效果越好。表 1 给出了四种聚类算法聚类结果的轮廓系数。

表 1 四种聚类算法聚类结果轮廓系数比较

聚类算法	轮廓系数	与本文算法比较
本文聚类算法	0.617	——
K-均值聚类算法	0.174	-0.443
K-中心点聚类算法	0.495	-0.122
FCM 聚类算法	0.223	-0.394

从表 1 可以看出,本文算法聚类结果的轮廓系数为0.617,分别比 K-均值聚类算法、K-中心点聚类算法、FCM 聚类算法高 0.443、0.122、0.394。因此,在聚类结果的类别数相同时,可以说明本文算法的聚类效果明显优于三种对比算法。

通过对四种算法聚类效果的定性分析和定量分析,进一步证明了最大偏差相似性准则和本文所提聚类算法在交通流曲线聚类中的优势和较好的实用性。

3.4 算法优点

本文提出的基于最大偏差相似性准则的交通流聚类算法具有如下优点:

a)本文采用的最大偏差相似性准则的原理简单、合理;最大偏差、相似度和偏离度的选择自由度使得最大偏差相似性准则对交通流曲线的相似性描述更加灵活,能够多精度度量交通流曲线的形态相似性;合适的偏离度能够避免交通流短时较大波动和数据噪声广泛分散的影响,客观合理,能满足实际应用的需要。

b)本文所提出的聚类算法无须预先指定类别数,避免了类别数的不正确设定可能导致聚类的不合理性;所提算法能够尽可能地将相似曲线聚到一类,保证所有类间曲线具有明显差异性和类内曲线具有高度的相似性,在聚类结果类别数相同时,其聚类效果明显优于 K-均值聚类算法、K-中心点聚类算法和 FCM 聚类算法。

4 结束语

本文针对交通流随机性强、波动频繁等特性,提出了一种新的交通流相似性度量准则——最大偏差相似性准则,并提出了基于最大偏差相似性准则的交通流聚类算法。所提出的最大偏差相似性准则能够较好地刻画频繁波动交通流曲线的形态相似性;最大偏差、相似度和偏离度的选择自由度保证了最大偏差相似性准则具有较广的使用范围、较大的应用灵活性。实验表明,所提出的聚类算法聚类效果较好,在类别数相同时,聚类效果比 K-均值聚类算法、K-中心点聚类算法、FCM 聚类算法更优、更实用,能够较好地满足实际应用的需要。最大偏差相似性准则参数的合适设置将有效提高本文聚类算法的聚类效果,增大其适用范围,今后将在不同实际应用中的最大偏差相似性准则参数设置方面进行深入的研究。

参考文献:

[1] Song Ying, Miller H J. Exploring traffic flow databases using space-time plots and data cubes[J]. Transportation, 2012, 39(2): 215-234.

[2] Habtemichael F G, Cetin M. Short-term traffic flow rate forecasting based on identifying similar traffic patterns[J]. Transportation Research, Part C: Emerging Technologies, 2016, 66(5): 61-78.

[3] Hou Yi, Edara P, Sun C. Traffic flow forecasting for urban work zones[J]. IEEE Trans on Intelligent Transportation Systems, 2015, 16(4): 1761-1770.

[4] Lam W H K, Wong S C, Lo H K. Emerging theories in traffic and transportation and methods for transportation planning and operations [J]. Transportation Research, Part C: Emerging Technologies, 2011, 19(2): 169-171.

(下转第 2292 页)

3.3.2 隐含因子个数的比较

K 的取值越大，RMSE 越小，表明在一定的范围内，随着隐含因子个数的增加，RMSE 值不断降低，推荐系统的准确度也就越高。

但由于隐含因子个数越大，分解后的矩阵维度越高，矩阵迭代分解的时间和空间复杂度越高，同时，隐含因子个数大到一定程度后 RMSE 趋于稳定。因此，不能通过简单增加隐含因子个数的办法来提高推荐系统的准确度。隐含因子个数的取值要综合考虑算法的复杂度与推荐准确度。

3.3.3 与其他推荐算法的比较

通过与其他两种模型结果的对比，本文提出的基于用户分类的隐含因子模型与其他两种模型相比 RMSE 值均降低了，这说明将用户分类信息融入到传统隐含因子模型中有助于提高算法预测的准确性。

4 结束语

本文通过在隐含因子模型中引入用户分类信息，提出基于用户分类的隐含因子模型。首先，该模型将推荐问题转换为优化问题，并将用户分类信息融入到矩阵迭代分解的过程中，弥补了传统的隐含因子模型忽略用户分类信息的缺陷；其次，通过引入用户的分类信息，使得在新用户加入推荐系统时，可以通过新用户的相关信息确定其所属的用户类别，从而为其提供个性化推荐服务，解决了新用户的冷启动问题；最后，通过与 LFM 和 USPMF 等模型在两个不同规模数据集上的对比实验，表明该模型不仅具有良好的稳定性，还能够有效提高预测推荐的准确性。

由于用户分类信息的加入，使得新模型产生了一些新的问题，如：a）该模型在原有模型的基础上加入了对分类矩阵的处理，增加了算法的复杂度，模型训练的时间较长；b）分类矩阵中的评分数据容易出现两极化，即使对奇异样本数据进行归一化处理，但对推荐结果仍然会产生一定的消极影响。这些问题将在今后的工作中作进一步的研究。

参考文献：

[1] Liu Hongxia. A survey of collaborative filtering technique in recommendation system[J]. Information Security & Technology, 2016, 110(4): 31-36.

[2] Guo Yanhong, Cheng Xuefen, Dong Dahai, *et al*. An improved collaborative filtering algorithm based on trust in e-commerce recommendation systems[C]//Proc of International Conference on Management and Service Science. New York: ACM Press, 2010: 1-4.

[3] Ghazarian S, Nematbakhsh M A. Enhancing memory-based collaborative filtering for group recommender systems[J]. Expert Systems with Applications, 2015, 42(7): 3801-3812.

[4] Lozano E, Gracia J, Collarana D, *et al*. Model-based and memory-based collaborative filtering algorithms for complex knowledge models [R]. Amsterdam: University of Amsterdam, Human Computer Studies Laboratory, 2011.

[5] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems[J]. Computer, 2009, 42(8): 30-37.

[6] Liu Juntao, Wu Caihua, Liu Wenyu. Bayesian probabilistic matrix factorization with social relations and item contents for recommendation [J]. Decision Support Systems, 2013, 55(3): 838-850.

[7] 王建芳，张朋飞，刘永利. 基于改进带偏置概率矩阵分解算法的研究[J]. 计算机应用研究，2017，34(5)：1397-1400，1414.

[8] 刘慧婷，陈艳，肖慧慧. 基于用户偏好的矩阵分解推荐算法[J]. 计算机应用，2015，35(S2)：118-121.

[9] 顾晔，吕红兵. 改进的增量奇异值分解协同过滤算法[J]. 计算机工程与应用，2011，47(11)：152-154.

[10] Funk S. Netflix update: try this at home[EB/OL]. 2006. http://sifter.org/? simon/journal/20061211.html.

[11] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]//Proc of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2008: 426-434.

[12] Zeng Xiangxiang, Ding Ningxiang, Zou Quan. Latent factor model with heterogeneous similarity regularization for predicting gene-disease associations[C]//Proc of IEEE International Conference on Bioinformatics and Biomedicine, Computer Society. New York: ACM Press, 2016: 682-687.

[13] 李卫平，杨杰. 基于随机梯度矩阵分解的社会网络推荐算法[J]. 计算机应用研究，2014，31(6)：1654-1656.

[14] Zhang Feng, Gong Ti, Lee V E, *et al*. Fast algorithms to evaluate collaborative filtering recommender systems[J]. Knowledge-Based Systems, 2016, 96(3): 96-103.

[15] Patrous Z S, Najafi S. Evaluating prediction accuracy for collaborative filtering algorithms in recommender system[J]. Information Sciences, 2016, 199(15): 78-92.

（上接第 2276 页）

[5] Yildirimoglu M, Geroliminis N. Experienced travel time prediction for congested freeways[J]. Transportation Research, Part B: Methodological, 2013, 53(4): 45-63.

[6] Salamanis A, Meladianos P, Kehagias D, *et al*. Evaluating the effect of time series segmentation on STARIMA-based traffic prediction model[C]//Proc of the 18th IEEE International Conference on Intelligent Transportation Systems. Piscataway, NJ: IEEE Press, 2015: 2225-2230.

[7] Nath R P D, Lee H J, Chowdhury N K, *et al*. Modified K-means clustering for travel time prediction based on historical traffic data [C]//Proc of the 14th IEEE International Conference on Knowledge-Based and Intelligent Information and Engineering Systems. Berlin: Springer, 2010: 511-521.

[8] Zhang Ting, Xia Yingjie, Zhu Qianqian, *et al*. Mining related information of traffic flows on lanes by K-medoids[C]//Proc of the 11th IEEE International Conference on Fuzzy Systems and Knowledge Discovery. Piscataway, NJ: IEEE Press, 2014: 390-396.

[9] Lin Lei, Wang Qian, Sadek A W. A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction[J]. Transportation Research, Part C: Emerging Technologies, 2015, 55(6): 444-459.

[10] Chen Ying, Kim J, Mahmassani H S. Pattern recognition using clustering algorithm for scenario definition in traffic simulation-based decision support systems[C]//Proc of the 17th IEEE International Conference on Intelligent Transportation Systems. Piscataway, NJ: IEEE Press, 2014: 798-803.

[11] Ma Xiaolei, Wu Y J, Wang Yinhai, *et al*. Mining smart card data for transit riders' travel patterns[J]. Transportation Research, Part C Emerging Technologies, 2013, 36(11): 1-12.

[12] Mei Yu, Tang Keshuang, Li Keping. Real-time identification of probe vehicle trajectories in the mixed traffic corridor[J]. Transportation Research, Part C: Emerging Technologies, 2015, 57(8): 55-67.

[13] Ding Yi, Fu Xian. Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm[J]. Neurocomputing, 2015, 188(5): 233-238.

[14] Lyu Yinghua, Ma Tinghuai, Tang Meili, *et al*. An efficient and scalable density-based clustering algorithm for datasets with complex structures[J]. Neurocomputing, 2015, 171(1): 9-22.

[15] Huang He, Tang Qifeng, Liu Zhen. Adaptive correction forecasting approach for urban traffic flow based on fuzzy-mean clustering and advanced neural network[J]. Journal of Applied Mathematics, 2013, 2013(1): 1-7.

[16] 李清泉，曹晶，乐阳，等. 短时车流量模式提取及时变特征分析[J]. 武汉大学学报：信息科学版，2011，36(12)：1392-1396.

[17] 闫伟，刘云岗，王桂华，等. 基于数据挖掘的交通流预测模型[J]. 系统工程理论与实践，2010，30(7)：1320-1325.

[18] Jiang Shan, Wang Shuofeng, Li Zhiheng, *et al*. Fluctuation similarity modeling for traffic flow time series: a clustering approach[C]//Proc of the 18th IEEE International Conference on Intelligent Transportation Systems. Piscataway, NJ: IEEE Press, 2015: 848-853.