

基于用户偏好特征挖掘的个性化推荐算法



刘晓飞 朱 斐 伏玉琛 刘 全

苏州大学计算机科学与技术学院 江苏 苏州 215006

(LXF_liuxf@163.com)

摘 要 为了提升社交网络个性化推荐能力,结合用户行为分布进行个性化推荐设计,文中提出基于用户行为特征挖掘的个性化推荐算法,构建社交网络的用户行为信息特征挖掘模型,采用显著数据分块检测方法对社交网络用户特征的行为信息进行融合处理,提取反映用户偏好的语义信息特征量。从情感、关键词和结构等方面根据用户行为特征组,结合模糊信息感知方法进行社交网络个性化推荐过程中的信息融合处理,在关联规则约束控制下,构建社交网络用户偏好特征的混合推荐模型,实现用户偏好特征挖掘,根据语义分布和用户的行为偏好实现社交网络的个性化信息推荐。仿真结果表明,采用所提方法进行社交网络个性化推荐的特征分辨能力较好,对用户行为特征的准确识别能力较强,提高了社交网络推荐输出的准确性。

关键词: 用户偏好;特征挖掘;个性化推荐;社交网络

中图法分类号 TP391

Personalized Recommendation Algorithm Based on User Preference Feature Mining

LIU Xiao-fei, ZHU Fei, FU Yu-chen and LIU Quan

School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

Abstract For the purpose of personalized recommendation ability of social network, this paper proposed a personalized recommendation algorithm based on user behavior feature mining according to the distribution of user behavior. The user behavior information feature mining model of social network is constructed, the big data fusion scheduling method is used to fuse the behavior information of social network user characteristics, and the semantic information features that reflect the user preference are extracted. According to the user's behavior feature groups from the aspects of emotion, keywords and structure, combined with the fuzzy information perception method, the information scheduling in the process of personalized recommendation of social network is carried out. Under the control of association rules constraints, a hybrid recommendation model of user preference features is constructed to realize user preference feature mining, and personalized information recommendation of social networks is realized according to semantic distribution and user behavior preference. The simulation results show that the proposed method has good feature resolution ability and accurate recognition ability to user behavior features, which improves the confidence level of social network recommendation output.

Keywords User preference, Feature mining, Personalized recommendation, Social network

随着网络自媒体的快速发展,大量的社交网络应运而生。社交网络是为了满足人们个性化需求而发展的自媒体网络模式。由于社交网络媒体的结构复杂,因此采用多分布的异构组网模式来完成社交网络模式的构建。随着网络上用户和商品数据的急剧增加,在社交网络中使用各种社交关系进行社交网络的个性化特征推荐,构建社交网络个性化推荐模型,结合特征提取和大数据挖掘方法,以实现社交网络推荐过程中的信息挖掘^[1],提取反映用户行为特征的社交网络个性化特征量,提高社交网络的信息检索和调度能力,极为必要。研究社交网络的个性化推荐方法在社交网络的信息化建设中具有重要意义,相关的推荐算法研究受到了人们的极大关注。

对社交网络的个性化推荐是建立在对社交网络的用户行为信息进行挖掘的基础上的,文献[2-3]通过提取社交网络的

统计信息特征量,结合模糊关联规则调度方法,进行社交网络推荐过程中的信息采样和模式识别。对社交网络进行推荐的传统方法主要有关联维特征提取方法、定量递归分析方法、模糊PID方法等^[2-3]。文献[4]提出的基于关联规则调度的社交网络个性化推荐算法,采用自适应学习方法进行社交网络推荐过程中的自适应寻优;但该方法进行社交网络推荐的计算开销较大,实时性不好。文献[5]提出的基于并行推荐的社交网络社区发现算法,根据并行推荐算法进行社交网络中的社区发现和优化推荐;但该方法进行社交网络推荐的寻优能力弱,收敛性不强。文献[6]提出的基于标签定位和并行挖掘的社交网络优化推荐算法,通过对社区属性特征进行挖掘实现社交网络的推荐;但该方法的特征匹配能力不好。文献[7]构建的基于用户行为挖掘的融合社交网络推荐模型,采用模

收稿日期:2019-07-25 返修日期:2019-11-16 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金项目(60673092)

This work was supported by the National Natural Science Foundation of China (60673092).

通信作者:伏玉琛(fuyuchen@163.com)

糊 C 均值聚类方法进行推荐信息特征的识别,根据聚类结果实现社区发现和社交网络推荐;该方法同样存在特征辨识能力不强和自适应性差的问题。

针对上述问题,本文提出了基于用户行为特征挖掘的个性化推荐算法。首先,构建社交网络的用户行为信息特征挖掘模型;然后,提取社交网络的用户特性化偏好特征量,并根据特征提取结果进行社交网络的用户行为偏好信息推荐,实现社交网络的个性化推荐优化;最后,通过仿真实验分析,展示了本文方法在提高社交网络个性化推荐能力方面的优越性。

1 社交网络个性化推荐的数据分布及特征分析

1.1 社交网络个性化推荐的数据分布

为了实现大数据环境下的分布式社交网络的个性化推荐,需要在异构存储结构模式下结合用户偏好信息,采用个性化信息融合方法,进行推荐模型构建^[8]。采用标签识别方法对社交网络个性化推荐进行信息采样,输出社交网络个性化推荐的状态特征量 $\mathbf{x}_j = \{x_{1j}, x_{2j}, \dots, x_{mj}\}^T$,以 2 倍以上的波特率进行社交网络的用户行为信息采样,得到社交网络个性化推荐的用户行为信息融合特征量为 $p(x_0)$ 。对采集的社交网络个性化推荐的用户行为信息进行联合特征挖掘,输出个性化推荐和联合特征采样模型为:

$$P_{ij}(k) = \frac{(l_j(k) - l_i(k))\eta_{ij}(k)}{\sum_{j \in N_i(k)} (l_j(k) - l_i(k))\eta_{ij}(k)} \quad (1)$$

采用相关性检测方法对社交网络个性化推荐特征信息进行联合识别,根据关联规则调度模型,得到社交网络个性化信息采样的样本集为 $s(t)$ 。利用物联网技术控制社交网络个性化推荐的过程,输出网络用户偏好特征样本集为:

$$s(v) = \int_0^v \sin\left(\frac{\pi}{2}x^2\right) dx \quad (2)$$

$$y(t) = u(s(t-\tau)) \exp(j\omega_c s(t-\tau)) \quad (3)$$

其中, v 表示社交网络个性化推荐的关联指向性系数; $u(t)$ 为网络用户偏好特征的有效信息特征分量,在异构分布式组网体系平台下,网络用户偏好特征分布的属性维数为 n ,即 N_1, \dots, N_n 。在 L 个时间片内,平台将数据聚集到 v_s 。社交网络个性化推荐信息的标准量化集为:

$$E_{Tx}(l, d) = E_{(Tx\text{-elec})}(l) + E_{(Tx\text{-amp})}(l, d) = lE_{\text{elec}} + l\epsilon d^n \\ = \begin{cases} lE_{\text{elec}} + l\epsilon_{fs}d^2, & d < d_0 \\ lE_{\text{elec}} + l\epsilon_{mp}d^4, & d \geq d_0 \end{cases} \quad (4)$$

文献[9]采用模糊自适应调度方法构建社交网络个性化推荐的大数据信息融合模型,并根据社交网络个性化推荐的数据分布特性进行个性化推荐算法的设计。

1.2 用户行为特征分析

文献[10]在构建社交网络的用户行为信息特征挖掘模型的基础上,采用显著数据分块检测方法进行社交网络用户特征的行为信息融合处理^[10],得到社交网络个性化推荐的资源调度函数为:

$$\hat{f}_i(n) = \frac{1}{2\pi i} \sum_{i=0}^n ia_i n^{i-1} \quad (5)$$

对提取的社交网络个性化推荐统计数据信息进行信息融合,在资源调度中心,根据用户偏好特征量进行多维特征重构^[11],采用多块融合匹配方法得到社交网络个性化推

荐的关联规则项为:

$$c_k = \frac{1}{j} \left[\frac{d^k}{d\omega^k} \ln(\Phi(\omega)) \right]_{\omega=0} = (-j)^k \left[\frac{d^k}{d\omega^k} \Psi(\omega) \right]_{\omega=0} \\ = (-j)^k \Psi^k(0) \quad (6)$$

对采集的社交网络个性化推荐资源信息进行自适应信息融合处理,构建底层数据库。数据融合聚类的最大分布间隔 $r_{\max} = (P(N_0\beta)^{-1})^{1/a}$,社交网络个性化推荐的样本空间分布距离 $d \leq r_{\max}$,社交网络个性化推荐的信息融合结果为:

$$X_a(m) = A_a \cdot e^{(j/2)\cot a \cdot m^2 \Delta a^2} \times \sum_{m=0}^{N-1} e^{(j/2)\cot a \cdot n^2 \Delta a^2} \\ e^{-j \frac{\text{sgn}(\sin a) \cdot 2\pi mn}{M}} \quad (7)$$

设计 ZigBee 数据采集节点作为社交网络个性化推荐的信息采样中心,采用联合关联规则挖掘技术提取社交网络个性化推荐的统计特征量,得到网络个性化推荐的关联规则集 $S = \{s_i | i = 1, 2, \dots, N_s\} (S \subset C)$ 。利用模糊相关性检测技术进行社交网络个性化推荐过程中的信息融合和去重处理,得到用户偏好特征的结构重组输出为:

$$S_p(u) = \{F^p[s(t)]\}(u) = \int_{-\infty}^{\infty} K_p(t, u) s(t) dt \quad (8)$$

采用 RFID 标签识别方法进行社交网络个性化推荐的信息采样,得到离散特征分布序列 x_1, x_2, \dots, x_{m+1} ,推荐信息输出的窗口函数为:

$$x_1 + x_2 + \dots + x_{m+1} = T + t - m \times t \quad (9)$$

采用自适应学习方法进行推荐过程中的优化分配,得到用户行为特征的优化分配模型为:

$$H_i(x) = \sum_{k=1}^K p_k \ln \frac{1}{p_k} = - \sum_{k=1}^K p_k \ln p_k \quad (10)$$

其中, K 表示社交网络个性化推荐的标签属性分布节点, p_k 为数据链路层的空间状态特征量。

构建社交网络个性化推荐的用户行为特征提取模型的过程如上所示。个性化推荐优化工作需要利用到特征提取结果^[12]。

2 社交网络个性化推荐算法的优化

2.1 网络用户行为偏好特征挖掘

在完成上述社交网络的用户行为信息特征挖掘模型的构建后,采用显著数据分块检测方法进行社交网络用户特征的行为信息融合处理,并实现社交网络的个性化推荐。在社交网络的驱动模块中,社交网络个性化推荐信息融合模型表达为:

$$\begin{cases} \mathbf{x} = (x_1, x_2, \dots, x_n) \\ \mathbf{y} = F(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x}))^T \end{cases} \quad (11)$$

其中, $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 为社交网络个性化推荐的用户行为特征分布集合; $\mathbf{y} = F(\mathbf{x})$ 表示存储空间分配节点集。

对不同领域的相关特征进行特征对齐,得到用户行为特征分布集合 $P(n_i) = \{p_k | pr_{kj} = 1, k = 1, 2, \dots, m\}$ 。采用结构一致学习算法^[13],得到空间链路分布集为:

$$RTT_s = (1-\alpha) \times RTT_s + \alpha \times RTT \quad (12)$$

根据用户行为类标信息和词性进行标签信息识别,得到个性化推荐的空间聚类模糊度函数 \bar{Q} 的优化值为:

$$Opti = \sum_{k=1}^m \alpha_i^{-1} \alpha_{it,k} \alpha_k^{-1} \alpha_j t_{k,j} \\ = \sum_{k=1}^m t_{i,k} t_{k,j} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases} \quad (13)$$

其中, α_k 为语义对齐二元特征组, α_j 为模糊修正向量, $t_{k,j}$ 为

用户行为偏好信息采样的时间间隔。

对非枢纽特征进行对齐处理,构建网络用户行为偏好特征挖掘模型,输出为:

$$P(\mathbf{U}|\alpha_U) = \prod_{i=1}^M N(U_i | 0, \alpha_U^{-1} \mathbf{I}) \quad (14)$$

$$P(\mathbf{V}|\alpha_V) = \prod_{j=1}^N N(V_j | 0, \alpha_V^{-1} \mathbf{I})$$

其中, $\alpha, \mathbf{U}, \mathbf{V}, \mathbf{I}$ 都是向量, \mathbf{U}, \mathbf{V} 为聚类特征向量。

综上分析,提取反映用户偏好的语义信息特征量,采用模糊信息感知技术进行社交网络个性化推荐过程中的信息融合和用户行为偏好特征挖掘^[14]。

2.2 社交网络的个性化联合推荐

假设分布式社交网络中用户偏好特征信息存储的节点的属性集为 $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$, 整体权重影响的计算方式为:

$$W_k(\mathbf{U}) = g \left(\frac{1}{M} \sum_{i=1}^M \frac{\sum_{j \in \text{Item}_i} r_{i,j,c_k}}{\sum_{j \in \text{Item}_i} s \times r_{i,j} + O} \right) \quad (15)$$

$$W_k(\mathbf{V}) = g \left(\frac{1}{M} \sum_{i=1}^N \frac{\sum_{j \in \text{User}_i} r_{i,j,c_k}}{\sum_{j \in \text{User}_i} s \times r_{i,j} + O} \right) \quad (16)$$

根据用户、项目的影响联合分布进行社交网络用户偏好特征的挖掘^[15], 构建矩阵分解模型, 得到社交网络的个性化联合特征分布满足:

$$\max_{x_{a,b,d,p}} \sum_{a \in A} \sum_{b \in B} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} V_p \quad (17)$$

$$\text{s. t. } \sum_{a \in A} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} R_p^{bw} \leq K_b^{bw}(S), b \in B \quad (18)$$

从情感、关键词和结构等方面, 根据用户行为特征组, 进行分布式用户偏好特征动态重构^[16]。根据数据的属性特征进行模糊聚类, 得到用户行为特征的信息识别概率为 $P(1) = [1 - L^{-1}]^{m-1}$ 。采用深度学习方法进行个性化推荐过程中的用户行为特征挖掘, 输出为:

$$P(k) = P(1) [1 - p(1)]^{k-1} \quad (19)$$

将邻域依赖特征投射到社交网络个性化推荐的模糊域中, 得到个性化特征模糊搜索均值函数为:

$$E(k) = \sum_{k=0}^{\infty} [1 - P(1)]^k = \frac{1}{1 - [1 - P(1)]} = \frac{1}{P(1)} \\ = \frac{1}{(1 - \frac{1}{L})^{m-1}} \quad (20)$$

个性化推荐的平均时隙数为:

$$T_{\text{Eary}} = E(k)L = \frac{L}{(1 - \frac{1}{L})^{m-1}} \quad (21)$$

计算社交网络的用户行为的联合信息熵, 得到个性化演化特征量。根据用户偏好特征挖掘结果^[17], 得到社交网络个性化推荐的联合分布密度函数描述为:

$$f_{T_1, T_2, \dots, T_n}(t_1, t_2, \dots, t_n) \\ = c(F_{T_1}(t_1), F_{T_2}(t_2), \dots, F_{T_n}(t_n)) \prod_{i=1}^n f_{T_i}(t_i) \\ = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{F}_T(\mathbf{t}) - \boldsymbol{\mu})' \Sigma^{-1}(\mathbf{F}_T(\mathbf{t}) - \boldsymbol{\mu})\right] \prod_{i=1}^n f_{T_i}(t_i) \quad (22)$$

以上是实现社交网络个性化推荐算法的优化全过程^[18-20]。

3 仿真实验与分析

为了测试本文方法在实现社交网络个性化推荐中的应用

性能, 结合 Matlab 进行仿真实验分析。社交网络推荐模型的数据来自云组合数据库 Pearson Database。对用户偏好特征挖掘的测试样本长度为 1024, 测试集规模为 2000, 用户行为特征分布属性的维数为 4, 社交网络个性化推荐的训练样本规模为 100, 信息采样频率为 80 kHz。用户行为偏好特征数据的分布如表 1 所列。

表 1 用户行为偏好特征数据的分布

Table 1 Distribution of user behavior preference characteristic data

数据集	统计值		回归分析值	
	测试集	训练集	测试集	训练集
汽车类数据集	2765	165	1643	465
美食类数据集	1567	546	1567	563
旅游类数据集	6465	67	4435	456
交通类数据集	4567	643	7567	378
电子信息类数据集	5655	46	6754	564
Book 数据集	6643	767	567	235

根据上述仿真环境和参数设定进行社交网络个性化推荐, 得到原始数据分布, 如图 1 所示。

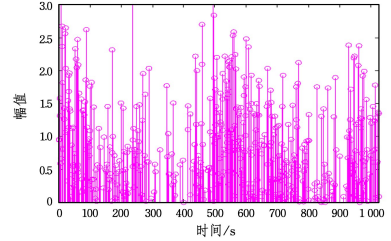


图 1 社交网络个性化推荐的原始数据采样

Fig. 1 Sampling of raw data recommended by personalization of social networks

以图 1 所示数据为测试对象, 提取反映用户偏好的语义信息特征量, 根据情感、关键词和结构等方面的信息, 进行社交网络中用户偏好信息的推荐, 构建用户行为特征组, 得到特征提取输出, 如图 2 所示。

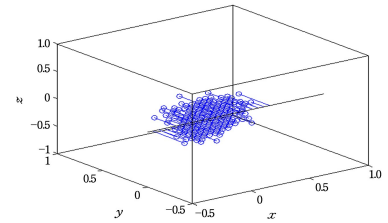


图 2 特征提取输出

Fig. 2 Feature extraction output

分析图 2 得知, 本文方法在进行社交网络个性化推荐时, 对用户行为特征的融合性能较好。在图 2 特征提取方法的基础上, 构建社交网络用户偏好特征的混合推荐模型, 实现用户偏好特征的挖掘, 得到个性化推荐的分类输出, 如图 3 所示。

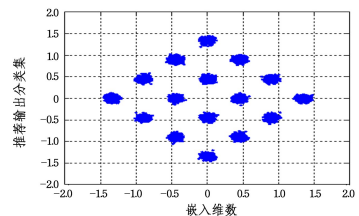


图 3 社交网络个性化推荐的分类输出

Fig. 3 Classified output of personalized recommendation for social networks

分析图3得知,采用本文方法进行推荐,对用户行为特征的准确识别能力较强。测试不同方法进行社交网络个性化推荐的准确性,得到结果如图4所示。分析图4得知,本文方法进行社交网络个性化推荐的精度较高。

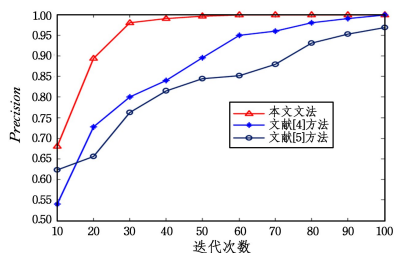


图4 社交网络个性化推荐的置信度对比

Fig. 4 Confidence comparison of personalized recommendation for social networks

结束语 本文结合特征提取和大数据挖掘方法实现社交网络推荐过程中的信息挖掘,构建社交网络个性化推荐模型,提出基于用户行为特征挖掘的个性化推荐算法。该算法采用显著数据分块检测方法进行社交网络用户特征的行为信息融合处理,根据用户偏好特征量进行多维特征重构,根据用户行为为类标信息和词性进行标签信息识别;从情感、关键词和结构等方面,根据用户行为特征组进行分布式用户偏好特征的动态重构,实现用户偏好特征挖掘,根据语义分布和用户的行为偏好实现社交网络的个性化信息推荐。研究得知,本文方法进行社交网络个性化推荐的信息融合性能较好,特征分辨聚类能力较强,置信度水平较高,在社交网络的信息推荐和网络设计中具有很好的应用价值。

参考文献

- [1] LI Y Z, ZHU Y Y, ZHONG M. k-core filtered influence maximization algorithms in social networks[J]. Journal of Computer Applications, 2018, 38(2): 464-470.
- [2] CAO J X, DONG D, XU S, et al. A k-core based algorithm for influence maximization in social networks[J]. Chinese Journal of Computers, 2015, 38(2): 238-248.
- [3] WANG L, MENG X F, GUO S N. Preservation of implicit privacy in spatio-temporal data publication[J]. Journal of Software, 2016, 27(8): 1922-1933.
- [4] YANG F R, ZHENG Y J, ZHANG C. Hybrid recommendation algorithm based on probability matrix factorization[J]. Journal of Computer Applications, 2018, 38(3): 644-649.
- [5] XIAO Y P, SUN H C, DAI T J, et al. A social network recommendation system scoring prediction method based on cloud model [J]. Journal of Electronics, 2018, 46(7): 1762-1767.
- [6] BI S, HO C K, ZHANG R. Wireless powered communication: opportunities and challenges[J]. IEEE Communications Magazine, 2015, 53(4): 117-125.
- [7] ULUKUS S, YENER A, ERKIP E, et al. Energy harvesting wireless communications; a review of recent advances[J]. IEEE Journal on Selected Areas in Communications, 2015, 33(3): 360-381.
- [8] ZHAO N, ZHANG S, YU R, et al. Exploiting interference for energy harvesting: a survey, research issues and challenges[J]. IEEE Access, 2017(5): 10403-10421.
- [9] CHEN W Z, ZHANG S, WANG D J, et al. A recommendation algorithm for elective courses in university based on nearest neighbor model and probability matrix factorization[J]. Journal of Liaoning Technical University, 2017(9): 976-982.
- [10] JIANG Y, ZHANG D F, DIAO Z L. Similarity Personalized Recommendation of User Matrix Model Based on Click Stream[J]. Computer Engineering, 2018, 44(1): 219-225.
- [11] CHEN H L. A Personalized Recommendation Algorithm Based on the Fusion of Trust Relation and Time Series[C]//IEEE International Conference on Computational Science & Engineering. IEEE, 2017.
- [12] SUN H, MA Y, YANG H B, et al. Collaborative Filtering Recommendation Algorithm by Optimizing Similarity and Clustering Users [J]. Journal of Chinese Computer Systems, 2014, 35(9): 1967-1970.
- [13] HU Y, JI B F, HUANG Y M, et al. Energy-efficient resource allocation algorithm for massive MIMO OFDMA downlink system [J]. Journal on Communications, 2015, 36(7): 40-47.
- [14] HU J, ZHU H W, MAO Y M. DBSCAN Clustering Algorithm Based on Adaptive Bee Colony Optimization[J]. Computer Engineering and Applications, 2019, 55(14): 105-114.
- [15] ZHAO Z B, SHI Y X, LI B Y. Newly-emerging Domain Word Detection Method Based on Syntactic Analysis and Term Vector [J]. Computer Science, 2019, 46(6): 29-34.
- [16] YANG J, WEI C H. Testing Serial Correlation in Partially Linear Additive Models[J]. Acta Mathematicae Applicatae Sinica, English Serie, 2019, 35(2): 401-411.
- [17] DOU Q, CHEN H, YU L Q, et al. Multi-level contextual 3D CNNs for false positive reduction in pulmonary nodule detection [J]. IEEE Transactions on Biomedical Engineering, 2017, 64(7): 1558-1567.
- [18] LEE G M, LEE J H. On nonsmooth optimality theorems for robust multiobjective optimization problems[J]. Journal of Nonlinear and Convex Analysis, 2015, 16(10): 2039-2052.
- [19] LONG Z Y, CHEN Z G, XU C L. Social Network Friend Recommendation Algorithm Based on User Interaction[J]. Computer Engineering, 2019, 45(3): 132-137.
- [20] LI L, ZHANG H N, LI Z B, et al. Research of Collaborative Filtering Recommendation Algorithm Fusion User Attributes in Government Purchase[J]. Journal of Chongqing University of Technology (Natural Science), 2015, 29(1): 76-81.



LIU Xiao-fei, born in 1995, postgraduate. His main research interests include machine learning and intelligence information processing.



FU Yu-chen, born in 1968, Ph.D, professor, is a member of China Computer Federation. His main research interests include reinforcement learning and intelligence information processing.