Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Cauchy Matrix Factorization for Tag-based Social Image Retrieval

XIAOYU DU[1,*], QIULI LIU[1,2,*], ZECHAO LI[3],(Member, IEEE), ZHIGUANG QIN[1],(Member, IEEE), AND JINHUI TANG[3],(Senior Member, IEEE)

[*]Xiaoyu Du and Qiuli Liu are Co-first authors.
[1]University of Electronic Science and Technology of China, Chengdu, Sichuan, 610054, China
[2]Beijing Normal University, Beijing, 100875, China
[3]Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094, China

Corresponding author: Zechao Li (e-mail: zechao.li@njust.edu.cn).

This work was partially supported by the National Key Research and Development Program of China under Grant 2016YFB1001001, the National Natural Science Foundation of China (Grant No. 61732007 and U1611461).

ABSTRACT User-provided tags associated with social images are essential information for social image retrieval. Unfortunately, these tags are often imperfect to describe the visual contents, which severely degrades the performance of image retrieval. Tag relevance learning models are proposed to improve the descriptive powers of tags mostly based on the Gaussian noise assumption. However, the intrinsic probability distribution of the noise is unknown and other probability distributions may be much better. Towards this end, this paper investigates the applicable probability distributions of tag noise and proposes a novel Cauchy Matrix Factorization (CMF) method for tag-based image retrieval. The Cauchy probability distribution is robust to all kinds of noise and more suitable to model the tagging noise of social images. Therefore, we utilize Cauchy distribution to model noise under the matrix factorization framework. Besides, other five probability density functions, i.e., Gaussian, Laplacian, Poisson, Student-t and Logistic, are investigated to model noise of social tags. To evaluate the performance of different probability distributions, extensive experiments on two widely-used datasets are conducted and results show the robustness of CMF to noisy tags of social images.

INDEX TERMS Cauchy noise, image retrieval, matrix factorization, social tags, tag relevance learning.

I. INTRODUCTION

Recent years have witnessed the explosive growth of social images associated with user-provided tags, which often makes users difficult to find their desired images. It raises an urgent demand for effective image retrieval technologies. Fortunately, the user-provided tags can describe the semantic information of the visual contents to a certain extent, which is beneficial to the promotion of tag-based image retrieval. On the other hand, everything is a double-edged sword. The quality of these tags cannot be guaranteed due to the limitations of tagging time and domain knowledge of amateur users. As reported in [Kennedy et al.2006], only about half of tags can well describe the visual contents of images. That is, the user-provided tags in real-world are usually imperfect with noisy and incomplete tags. Therefore, it is necessary but challenging to improve the descriptive powers of tags with respect to the visual contents of social images.

Many methods have been proposed for image retagging and tag refinement by removing the noisy tags and complementing the relevant but missing tags [Li et al.2009], [Zhu et al.2010], [Liu et al.2010], [Znaidia et al.2013], [Wu et al.2013], [Niu et al.2015], [Xue et al.2016], [Li and Tang2017b], [Tang et al.2017], [Li and Tang2017a], [Zhang et al.2018]. Most of these methods learn the image-tag relevance by minimizing the prediction error based on the Matrix Factorization (MF) framework. The refined tagging matrix is learned by minimizing the tagging noise under the low-rank constraint and considering the content consistency and tag correlation in [Zhu et al.2010], [Li and Tang2017b]. In [Wu et al.2013], the relevance of images to tags is refined by constraining it to be consistent with the original one. The tagging matrix is reconstructed by two latent factor matrices, which are learned by minimizing the difference between the refined tagging matrix and the observed tagging matrix in [Tang et al.2017], [Li and Tang2017a]. The widely-used function is to minimize the sum of squared error with some regularization terms that help prevent models from overfitting.

The problem that minimizes the sum-of-squared-errors objective function under MF has the underlying assumption about Gaussian distribution [Salakhutdinov and Mnih2008], [Ma et al.2011]. Gaussian probability density function the most well-known and widely used distribution in many fields such as signal processing and image analysis due to its simplicity and convenient solving as well as the central limit theorem [Park et al.2013]. However, the intrinsic probability distribution of data in real-world is unknown. It may be unsuitable to model the real-world data. In the real world, there are various types of noise, such as large, small, dense and sparse noise [Xie and Xing2014]. It is well known that Gaussian-based methods are limited to small noise and sensitive to the noise of large magnitude. Consequently, it is necessary and important to propose a new MF method to deal with various types of noise for tag-based image retrieval.

Towards this end, this paper proposes a novel Cauchy Matrix Factorization (CMF) method by modeling the tagging noise using the Cauchy distribution. The tradition matrix factorization model is based on the Gaussian noise assumption, which leads to the minimizing optimization problem with the sum-of-squared-errors objective function. However, it is sensitive to large noise. To well model noise, we deeply investigate the applicable probability distributions based on different probabilistic noise assumptions, i.e., Laplacian, Poisson, Student-t, Logistic and Cauchy distributions, and find the Cauchy assumptions is robust to all kinds of noise. Therefore, the Cauchy distribution is explored to model noise and CMF is derived. It is formulated as an optimization problem with a well-defined objective function. And the corresponding simple yet efficient updating algorithm is developed. To evaluate the performance, extensive experiments are conducted on two widely-used social image benchmarks for tag-based social image retrieval.

II. RELATED WORKS

In this section, we will briefly review some previous methods on tag-based social image retrieval and matrix factorization.

A. TAG-BASED SOCIAL IMAGE RETRIEVAL

It is essential to learn the idea image-tag relevance for tag-based social image retrieval. It is related to the traditional image annotation [Barnard et al.2003], [Wong and Leung2008], [Makadia et al.2010], [Yao et al.2019], which learns models based on the supervised training data. With the popularity of social networks and intelligent devices, images associated weakly-supervised user-provided tags dramatically increase. To improve the performance of tag-based social image retrieval, many methods have been proposed to refine the image-tag relevance by exploring the weakly-supervised tagging information, including shallow methods [Li et al.2009], [Zhu et al.2010], [Liu et al.2010], [Znaidia et al.2013], [Wu et al.2013], [Feng et al.2014], [Niu et al.2015], [Xue et al.2016], [Tang et al.2017] and deep methods [Fu et al.2015], [Li and Tang2017b], [Nguyen et al.2017], [Li and Tang2017a], [Zhang et al.2018], [Li et al.2019].

For the shallow methods, they learn the desired tag relevance based on the conventional learning models by exploring the user-provided tag information. In [Li et al.2009], [Znaidia et al.2013], the image-tag relevance is refined by the neighbor voting strategy based on the hand-crafted visual features. In [Li et al.2009], each visual neighbor is treated equally. In [Liu et al.2010], the consistency between visual similarity and semantic similarity is explored to remove the imprecise tags and add the relevant tags. A low-rank matrix decomposition is introduced to address the tag refinement problem in [Zhu et al.2010]. Gong et al. [Gong et al.2013] proposed a model to explore the user-provided tag information by clustering them as topics. In [Wu et al.2013], the relevance between images and tags is learned by requiring it to be consistent with the observed one and exploring the visual similarity. The dual sparse reconstruction approach was proposed for social image tag completion in [Lin et al.2013]. The missing tags are complemented and the noisy tags are removed by minimizing the tagging noise under the low-rank matrix recovery framework in [Feng et al.2014]. In [Xue et al.2016], image tagging is performed with multi-view representation learning with the sum-of-squared-errors objective function. Factor analysis model has been explored to discover the tag relevance [Niu et al.2015], [Tang et al.2017]. The tensor factorization model is proposed to learn the image-tag relevance by minimizing with the sum-of-squared-errors objective function between the learned image-tag-user tensor and the observed one in [Tang et al.2017]. Most of the above methods learn the desired image-tag relevance matrix by minimizing the errors between it and the observed one with different constraints.

Recent years, the deep neural networks, such as Convolutional neural networks (CNNs) [Krizhevsky et al.2012], [Du et al.2017b], have been widely used due to its amazing performance in the visual-related applications [Du et al.2017a]. Consequently, many deep methods have been proposed to learn the image-tag relevance. Rather than directly using the CNN features, a deep nonnegative low-rank model is proposed to refine tags by jointly exploring the low-rank model and deep feature learning in [Li and Tang2017b]. In [Nguyen et al.2017], the tag relevance learning is performed by using deep transfer learning based on the tagging information and visual features. A deep matrix factorization model is proposed to refine tags of social images by jointly exploring the deep learning and local learning under the matrix factorization framework in [Li and Tang2017a]. In [Zhang et al.2018], the user-provided tags are refined by using the deep neural network as the image feature learning, as well as exploring visual consistency, semantic dependency, and user-error sparsity simultaneously. A unified deep collaborative embedding model is proposed by incorporating the deep learning and factor analysis for the optimal compatibility of representation learning and latent space discovery in [Li et al.2019]. It integrates the weakly-supervised tagging information, image similarity, and tag correlation simultaneously and seamlessly by the collective matrix factorization model. The sum-of-squared-errors objective function with the Gaussian noise assumption is formulated.

The aforementioned methods are usually based on the Gaussian noise assumption. Different from them, we propose to investigate the probability distributions for tag-based social image retrieval by introducing different probability distribution assumptions.

B. MATRIX FACTORIZATION

Matrix Factorization (MF) has been widely used in different application, which decomposes a matrix as the product of two factor matrices. The traditional matrix factorization model has the underlying Gaussian observation noise assumption as the probabilistic interpretation in Probabilistic Matrix Factorization (PMF) [Salakhutdinov and Mnih2008]. Some methods have been developed to improve the classic matrix factorization model [Lee and Seung1999], [Chiang et al.2015]. Nonnegative Matrix Factorization (NMF) [Lee and Seung1999] requires each element of the latent matrices to be nonnegative. Recently some deep matrix factorization models have been proposed to learn a hierarchy of hidden representations in [Li and Tang2017a], [Trigeorgis et al.2017]. Most of these matrix factorization methods are formulated with the sum-of-squared-errors objective function. The underlying assumption is the Gaussian noise, which may be unsuitable in real-world applications. It is necessary to investigate different probability distributions to formulate the matrix factorization models. PMoEP [Cao et al.2016] and LRMF-MoG [Dong et al.2017] leverage the mixture of distributions to model the noises, yet they also take in the complex structure. Therefore, in this work, we conduct the study to investigate different probability distributions under the matrix factorization framework for tag-based social image retrieval.

III. CAUCHY MATRIX FACTORIZATION

In this section, we will investigate different probability distribution assumptions and elaborate the proposed CMF method.

A. PRELIMINARIES

In this paper, we use bold uppercase characters and bold lowercase characters to denote matrices and vectors, respectively. The lowercase character is used to denote the scalar. For any matrix $\mathbf{A}$, $\mathbf{a}_i$ denotes its i-th column vector while $\mathbf{a}^j$ is its j-th row vector. The (i,j)-element of $\mathbf{A}$ is denoted by A_{ij} For matrix operation, $\mathbf{A}^T$ is the transposed matrix of $\mathbf{A}$, while $\mathrm{Tr}[\mathbf{A}]$ is the trace of $\mathbf{A}$ if $\mathbf{A}$ is square. The Frobenius norm of $\mathbf{A} \in \mathbb{R}^{m\times n}$ is defined as $\|\mathbf{A}\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n A_{ij}^2 = \mathrm{Tr}[\mathbf{A}^T\mathbf{A}]$. The ℓ_1-norm for $\mathbf{A}$ is defined as $\|\mathbf{A}\|_1 = \sum_{i=1}^m \|\mathbf{a}^i\| = \sum_{i=1}^m \sum_{j=1}^n |A_{ij}|$.

Considering a social image set, there are n images $\{\mathbf{x}_i\}_{i=1}^n$ associated with m user-provided tags $\mathcal{C} = \{t_1, t_2, \cdots, t_m\}$. For the i-th image, its observed relevance to tags is represented as a m-dimensional binary-valued vector $\{\mathbf{y}_i\}$. The tagging matrix is denoted as $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_n] \in \mathbb{R}^{m\times n}$. The i-th row vector of $\mathbf{Y}$ corresponds to a tagging vector of all the images with respect to the i-th tag. $Y_{ij} = 1$ indicates that $\mathbf{x}_j$ is associated with the i-th tag, and $Y_{ij} = 0$ otherwise. For the tag relevance learning, an ideal tagging matrix $\mathbf{F} \in \{0,1\}^{m\times n}$ is desired.

Matrix factorization is to decompose the observed matrix into two factor matrices $\mathbf{U} \in \mathbb{R}^{r\times m}$ and $\mathbf{V} \in \mathbb{R}^{r\times n}$. The i-th column of $\mathbf{U}$ is denoted as $\mathbf{u}_i$, while the j-th column of $\mathbf{V}$ is denoted as $\mathbf{v}_j$. The traditional matrix factorization model, such as Singular Value Decomposition (SVD), is formulated as follows,

$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2}\|\mathbf{Y} - \mathbf{U}^T\mathbf{V}\|_F^2 + \frac{\lambda_1}{2}\|\mathbf{U}\|_F^2 + \frac{\lambda_2}{2}\|\mathbf{V}\|_F^2. \tag{1}$$

The last two terms are introduced as the regularizers to avoid overfitting with two positive parameters λ_1 and λ_2. Thus the ideal tagging matrix $\mathbf{F}$ is obtained by $\mathbf{F} = \mathbf{U}^T\mathbf{V}$, where the element F_{ij} indicates whether the image $\mathbf{x}_j$ is associated with the tag t_i.

B. INVESTIGATION

In this subsection, we investigate the impact of the noise distribution hypothesis over matrix factorization. Since all of them aim to estimate the posterior probability $p(\mathbf{U},\mathbf{V}|\mathbf{Y})$. Through the simple Bayesian inference, the posterior distribution $p(\mathbf{U},\mathbf{V}|\mathbf{Y})$ of $\mathbf{U}$ and $\mathbf{V}$ given $\mathbf{Y}$ can be easily obtained by,

$$p(\mathbf{U},\mathbf{V}|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{U},\mathbf{V})p(\mathbf{U})p(\mathbf{V}). \tag{2}$$

The latent vectors $\mathbf{u}_i$ and $\mathbf{v}_j$ are assumed with zero-mean spherical Gaussian priors.

$$p(\mathbf{U}|\sigma_U^2) = \prod_{i=1}^m \mathcal{N}(\mathbf{u}_i|0, \sigma_U^2\mathbf{I}), \tag{3}$$

$$p(\mathbf{V}|\sigma_V^2) = \prod_{j=1}^n \mathcal{N}(\mathbf{v}_j|0, \sigma_V^2\mathbf{I}). \tag{4}$$

Then, the posterior distribution $p(\mathbf{U},\mathbf{V}|\mathbf{Y})$ of $\mathbf{U}$ and $\mathbf{V}$ given $\mathbf{Y}$ can be obtained. By maximizing the log of the posterior distribution, we have the following problem,

$$\max_{\mathbf{U},\mathbf{V}} \prod_{i=1}^m \prod_{j=1}^n p(Y_{ij}|F_{ij}) \prod_{k=1}^m p(\mathbf{u}_k) \prod_{l=1}^n p(\mathbf{v}_l), \tag{5}$$

which is equivalent to

$$\max_{\mathbf{U},\mathbf{V}} \sum_{i=1}^m \sum_{j=1}^n \ln p(Y_{ij}|F_{ij}) + \lambda_1\|\mathbf{U}\|_F^2 + \lambda_2\|\mathbf{V}\|_F^2. \tag{6}$$

The regularization terms $\lambda_1\|\mathbf{U}\|_F^2$ and $\lambda_2\|\mathbf{V}\|_F^2$ are brought in according to the log function over Gaussian prior, which is,

$$\sum_{k=1}^{m}\ln p(\mathbf{u}_k)+\sum_{l=1}^{n}\ln p(\mathbf{v}_l)$$
$$=\underbrace{-\frac{1}{2\sigma_U^2}\sum_{k=1}^{m}\mathbf{u}_k^T\mathbf{u}_k-\frac{1}{2\sigma_V^2}\sum_{l=1}^{n}\mathbf{v}_l^T\mathbf{v}_l}_{\text{Regularizations}}+\underbrace{C}_{\text{Constants}}. \tag{7}$$

Finally, according to Eq. (6), the main task is to seek the probability $p(Y_{ij}|F_{ij})$.

Gaussian. The traditional MF model in Eq. (1) has the Gaussian noise assumption [Salakhutdinov and Mnih2008]. It assumes that each elements E_{ij} of $\mathbf{E}=\mathbf{Y}-\mathbf{F}$ is subject to Gaussian distribution with mean 0 and variance σ^2.

$$p(E_{ij})=\mathcal{N}(E_{ij}|0,\sigma^2)=\mathcal{N}(Y_{ij}-\mathbf{u}_i^T\mathbf{v}_j|0,\sigma^2). \tag{8}$$

The conditional distribution over the observed values is obtained by,

$$p(\mathbf{Y}|\mathbf{U},\mathbf{V},\sigma^2)=\prod_{i=1}^{m}\prod_{j=1}^{n}\mathcal{N}(Y_{ij}|\mathbf{u}_i^T\mathbf{v}_j,\sigma^2). \tag{9}$$

According to Eq. (6), we can obtain the same optimization problem as the one in Eq. (1). The gradient descent algorithm can be used to find a local minimum. The Gaussian distribution may perform well with small noise, but it is sensitive to the noise of large magnitude.

Laplacian. For the social image tag refinement, there are some missing tags. That is, some elements of $\mathbf{Y}$ are unknown, which results in that the objective function with the Frobenius norm in Eq. (1) is not the most appropriate. The Laplacian distribution is often used to reduce sensitivity to the outliers in the data [Eriksson and van den Hengel2010]. The Laplacian distribution with mean 0 and scale b is used to model the noises and all the elements are independent.

$$p(Y_{ij}|F_{ij})=p(E_{ij})=\frac{1}{2b}\exp(-\frac{|E_{ij}|}{b}). \tag{10}$$

According to Eq. (6), we can have the following optimizing problem with the ℓ_1-norm loss function.

$$\min_{\mathbf{U},\mathbf{V}}\|\mathbf{Y}-\mathbf{U}^T\mathbf{V}\|_1+\lambda_1\|\mathbf{U}\|_F^2+\lambda_2\|\mathbf{V}\|_F^2. \tag{11}$$

It is well known that the ℓ_1 norm is more robust to outliers than the ℓ_2 norm. However, it does not enable to address the dense noise.

Poisson. It is known that the Poisson distribution is used to model the independent variables in whole numbers.

$$p(y)=\frac{\exp(-\lambda)\lambda^y}{y!} \tag{12}$$

Here λ is the mean value. Hence, by setting the mean value to F_{ij}, the Poisson distribution of Y_{ij} given F_{ij} is defined as

$$p(Y_{ij}|F_{ij})=\frac{\exp(-F_{ij})F_{ij}^{Y_{ij}}}{Y_{ij}!}. \tag{13}$$

According to Eq. (6), by maximizing the log of the posterior distribution, we have the following optimization problem.

$$\min_{\mathbf{U},\mathbf{V}}\sum_{i=1}^{m}\sum_{j=1}^{n}[\mathbf{u}_i^T\mathbf{v}_j-Y_{ij}\ln(\mathbf{u}_i^T\mathbf{v}_j)]$$
$$+\lambda_1\|\mathbf{U}\|_F^2+\lambda_2\|\mathbf{V}\|_F^2. \tag{14}$$

Student-t. The Student-t distribution is a heavy tailed generalization of the Gaussian distribution, which can handle atypical observations [Archambeau et al.2006].

$$p(y)=\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}(1+\frac{y^2}{\nu})^{-\frac{\nu+1}{2}}. \tag{15}$$

Here $\Gamma(\cdot)$ denotes the Gamma function and $\nu>0$ is the number of degrees of freedom. Hence, the Student-t distribution of Y_{ij} given F_{ij} is defined as

$$p(Y_{ij}|F_{ij})=\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}(1+\frac{(Y_{ij}-F_{ij})^2}{\nu})^{-\frac{\nu+1}{2}}. \tag{16}$$

According to Eq. (6), we can have the posterior distribution of $\mathbf{U}$ and $\mathbf{V}$ given $\mathbf{Y}$. The objective function based on the Student-t distribution is obtained by maximizing the log of the posterior distribution.

$$\min_{\mathbf{U},\mathbf{V}}\sum_{i=1}^{m}\sum_{j=1}^{n}\ln(1+\frac{(Y_{ij}-\mathbf{u}_i^T\mathbf{v}_j)^2}{\nu})$$
$$+\lambda_1\|\mathbf{U}\|_F^2+\lambda_2\|\mathbf{V}\|_F^2. \tag{17}$$

The Student-t distribution can avoid the disadvantages of the Laplacian distribution and the Gaussian distribution, but cannot well address the problem of large noises.

Logistic. The logistic distribution resembles the Gaussian distribution in shape but has heavier tails. The noise E_{ij} is modeled by utilizing the logistic distribution with mean 0.

$$p(E_{ij})=\frac{\exp(-\frac{E_{ij}}{b})}{b[1+\exp(-\frac{E_{ij}}{b})]^2}. \tag{18}$$

Here b is the scale parameter. According to Eq. (6), the posterior distribution $p(\mathbf{U},\mathbf{V}|\mathbf{Y})$ can be obtained by the simple Bayesian inference. By maximizing the log of the posterior distribution, the matrix factorization model with the logistic distribution can be formulated as the following minimizing problem.

$$\min_{\mathbf{U},\mathbf{V}}\sum_{i=1}^{m}\sum_{j=1}^{n}(\frac{Y_{ij}-\mathbf{u}_i^T\mathbf{v}_j}{b}+2\ln(1+\exp(-\frac{Y_{ij}-\mathbf{u}_i^T\mathbf{v}_j}{b})))$$
$$+\lambda_1\|\mathbf{U}\|_F^2+\lambda_2\|\mathbf{V}\|_F^2. \tag{19}$$

The behavior of the Logistic distribution in modeling noise is very similar to the Gaussian distribution.

C. CMF

Actually, there are many types of noise in the real-world data, such as sparse, dense, small as well as large [Xie and Xing2014]. And it is impossible to know the intrinsic probability distribution of data in the real world. The above distributions enable to deal with one kind of noise. For example, the Gaussian distribution is able to address the problem of small noise but sensitive to large noise. That is, the above noise assumptions may be unsuitable to model the real-world data. It is well known that the Cauchy distribution is smooth at the value of the location parameter, which makes it suitable to model the dense noise. Besides, it is capable of modeling the large noise due to its heavy tail [Xie and Xing2014]. That is, it has the ability to deal with various types of noise. Therefore, we propose to explore the Cauchy distribution for social tag noise modeling.

The noise E_{ij} is modeled by utilizing the Cauchy distribution with local parameter zero.

$$p(E_{ij}) = \frac{b}{\pi} \frac{1}{b^2 + E_{ij}^2}. \tag{20}$$

Here b is the scale parameter. The Cauchy distribution of Y_{ij} given F_{ij} is defined as,

$$p(Y_{ij}|F_{ij}) = \frac{b}{\pi} \frac{1}{b^2 + (Y_{ij} - F_{ij})^2}. \tag{21}$$

The Cauchy distribution of $\mathbf{Y}$ given $\mathbf{F}$ is defined as follows.

$$p(\mathbf{Y}|\mathbf{U},\mathbf{V}) = \prod_{i=1}^{m}\prod_{j=1}^{n} \frac{b}{\pi} \frac{1}{b^2 + (Y_{ij} - F_{ij})^2}. \tag{22}$$

According to Eq. (6), the posterior distribution $p(\mathbf{U},\mathbf{V}|\mathbf{Y})$ of $\mathbf{U}$ and $\mathbf{V}$ given $\mathbf{Y}$ can be obtained. By maximizing the log of the posterior distribution, we have the following problem.

$$\begin{aligned}\max_{\mathbf{U},\mathbf{V}} -\sum_{i=1}^{m}\sum_{j=1}^{n} \ln(b^2 + (Y_{ij} - \mathbf{u}_i^T\mathbf{v}_j)^2) \\ -(\lambda_1 \|\mathbf{U}\|_F^2 + \lambda_2 \|\mathbf{V}\|_F^2).\end{aligned} \tag{23}$$

The above maximizing problem is equivalent to the following minimizing problem.

$$\begin{aligned}\min_{\mathbf{U},\mathbf{V}} \sum_{i=1}^{m}\sum_{j=1}^{n} \ln(b^2 + (Y_{ij} - \mathbf{u}_i^T\mathbf{v}_j)^2) \\ +\lambda_1 \|\mathbf{U}\|_F^2 + \lambda_2 \|\mathbf{V}\|_F^2.\end{aligned} \tag{24}$$

The above problem can be efficiently solved by the gradient descent algorithm, which is an iterative process. Suppose L is the observed value of the loss function, in each iteration, we compute the gradients via $\partial L/\partial\mathbf{U}$ and $\partial L/\partial\mathbf{V}$, and update the parameters $\mathbf{U}$ and $\mathbf{V}$ with the gradients. After T iterations, the $\mathbf{U}$ and $\mathbf{V}$ are considered to be the optimal parameters. In our experiments, we leverage Adagrad [Duchi et al.2011] to update the parameters, and T is fixed to 2000.

TABLE 1: Statistics of the community-contributed datasets with image and tag counts in the format mean / maximum.

	MIRFlickr	NUS-WIDE
Tag size	457	3137
Concept size	18	81
Image size	25,000	269,648
Tags per image	2.7 / 45	7.9 / 201
Concepts per image	4.7 / 17	1.9 / 13
Images per tag	145.4 / 1,483	677.1 / 20,140
Images per concept	3,102.8 / 10,373	6,220.3 / 74,190

IV. EXPERIMENTS

In this section, we conduct experiments to evaluate the performance of the proposed Cauchy matrix factorization method for tag-based social image retrieval.

A. DATASET

Experiments are conducted on two social image datasets, i.e., MIRFlickr [Huiskes and Lew2008] and NUS-WIDE [Chua et al.2009], which have been widely used for social image understanding and retrieval tasks. Each image is associated with several user-provided tags.

The **MIRFlickr** dataset contains $25,000$ images associated with $1,386$ tags. Due to some obviously noisy tags, tags that appear less than 50 times are removed, resulting in a vocabulary of 457 tags. The ground-truth annotations of 18 concepts are preserved, which are used to evaluate the performance.

The **NUS-WIDE** dataset has $269,648$ images associated with $5,018$ tags. Due to some misspelt or meaningless tags, those tags whose occurrence numbers below 125 are removed. And we obtained $3,137$ unique tags. The ground-truth annotations of 81 concepts are also provided.

Some statistics of this social image dataset are summarized in Table 1.

B. SETTINGS

To evaluate the ranking order of tag-based image retrieval, experimental results with single-tag queries are analyzed. It is well known that Average Precision (AP) is the widely-used measure used for search. Mean Average Precision (mAP) is obtained by averaging AP over all the concepts. In experiments, mAP is used to evaluate the performance. Besides, the area under the receiver operating characteristic curve (AUC), has been used as the standard and more faithful measure for model comparison in many applications. Thus, AUC is taken into account as the evaluation metric. In experiments, the microaveraging and macroaveraging measures are introduced to evaluate the global performance across multiple concepts and the average performance of all the concepts.

For comparison, the original user-provided tags are adopted to calculate the results, which is used as the baseline. The matrix factorization methods with different probability distributions, denoted as GMF (Gaussian), LaMF (Laplacian), PoMF (Poisson), StMF (Student-t) and LoMF (Logistic), are compared to show the effectiveness of CMF. There are several parameters to be set in advance, such as the dimension r and the overfitting parameters λ_1 and λ_2. There are some other parameters, such as the degree of freedom ν for the Student-t distribution, the scale parameter b for the Cauchy distribution, as well as the scale parameter b for the logistic distribution. r is empirically set to 150 and 300 for MIRFlickr and NUS-WIDE, respectively. We set $\lambda_1 = \lambda_2 = 0.005$ empirically. For other parameters, the grid-search strategy is utilized over $\{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$. To alleviate the instability introduced by initialization, experiments are independently repeated 5 times, and the average values are reported.

C. RESULTS

We first carry out experiments to evaluate the performance for tag-based social image retrieval in terms of mAP. The results on the MIRFlickr and NUS-WIDE datasests are presented in Figure 1.

From the results, it can be observed that all the matrix factorization models improve the quality of tags and make the results of tag-based image search better. Second, CMF achieves the best retrieval performance, which demonstrates that the Cauchy probability distribution may be more suitable to model the tagging noise of social images. The Cauchy probability distribution can deal with several forms of noise patterns [Xie and Xing2014]. Third, LaMF is better than GMF. The Gaussian probability distribution is able to well model the small noise and the Laplacian probability distribution is suitable for the sparse noise. Thus, the noise of the user-provided tags is somewhat sparse. Forth, the performance of StMF is slightly superior to the performance of GMF, but worse than the performance of CMF and LaMF. Fifth, LoMF is just better than the original tagging and worse than other methods. Besides, PoMF achieves better results in terms of mAP than the original tags and LoMF, but worse results than other methods. It may be that it is somewhat more suitable to model the tagging noise than the logistic distribution. In a word, it may be suitable to construct models based on the Cauchy noise assumption for tag-based social image retrieval, and it is necessary and useful to conduct experiments to investigate the effectiveness of different noise assumptions.

Besides, additional experiments are conducted to compare the performance in terms of MicroAUC and MacroAUC, and the corresponding results are shown in Table 2 and Table 3. From the results, it can be seen that CMF gains the best results in terms of AUC, which is consistent with the observations in the above experiments. That is, the Cauchy distribution is appropriate for noise modeling in the tag-based image retrieval task. Besides, the performance is improved by dealing with the noise of the user-provided tags with these several probability distributions. That is, the models based on these probability distributions can reduce the tag noises to some extent. GMF also enables to address the tagging noise but performs worse than LaMF. Furthermore, the similar observations to the ones from the results in terms of mAP can be obviously seen. Finally, it is meaningful to adopt a suitable assumption to model the noise of social images.

TABLE 2: Experimental results (mean microauc ± standard deviation, mean macroauc± standard deviation) on the MIRFlickr dataset.

Method	MicroAUC	MacroAUC
Baseline	0.642 ± 0	0.634 ± 0
GMF	0.653 ± 0.005	0.646 ± 0.003
LaMF	0.664 ± 0.006	0.649 ± 0.004
PoMF	0.670 ± 0.003	0.654 ± 0.004
StMF	0.679 ± 0.006	0.663 ± 0.001
LoMF	0.654 ± 0.005	0.641 ± 0.003
CMF	**0.692 ± 0.002**	**0.667 ± 0.002**

TABLE 3: Experimental results (mean microauc ± standard deviation, mean macroauc± standard deviation) on the NUS-WIDE dataset.

Method	MicroAUC	MacroAUC
Baseline	0.752 ± 0	0.642 ± 0
GMF	0.769 ± 0.003	0.661 ± 0.002
LaMF	0.775 ± 0.005	0.710 ± 0.006
PoMF	0.772 ± 0.004	0.691 ± 0.002
StMF	0.779 ± 0.005	0.709 ± 0.003
LoMF	0.763 ± 0.003	0.677 ± 0.002
CMF	**0.785 ± 0.002**	**0.737 ± 0.001**

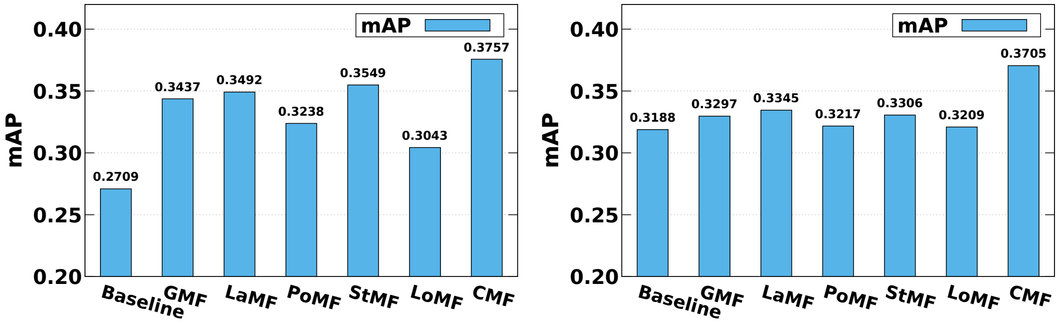
It is supposed to construct the learning model based on the appropriate probability distribution for the desired applications. Although the Gaussian distribution has been widely used, it may be unsuitable in the real-world tasks.

D. SENSITIVENESS ANALYSIS

The dimension r is an important hyper-parameter for matrix factorization based methods. In this section, experiments are conducted to evaluate the effect of r. The results in terms of mAP are presented in Figure 2.

It can be easily observed that r has somewhat effort on the performance of tag-based image search. And relatively, its values have great impacts on performance based on the Student-t probability distribution. If the value of r is set to a small value, the performance becomes poor, which is even worse than the performance of the original user-provided tags. When the value of the parameter r becomes larger, the results of the matrix factorization models with different probability distributions become better. But the computational cost increases with large value of the parameter r. By comprehensively considering the effectiveness and the computational cost, we set r to 150 and 300 for all the compared methods on the MIRFlick and NUS-WIDE datasets in experiments, respectively. How to adaptively identify the value of the parameter r will be studied in our future work.

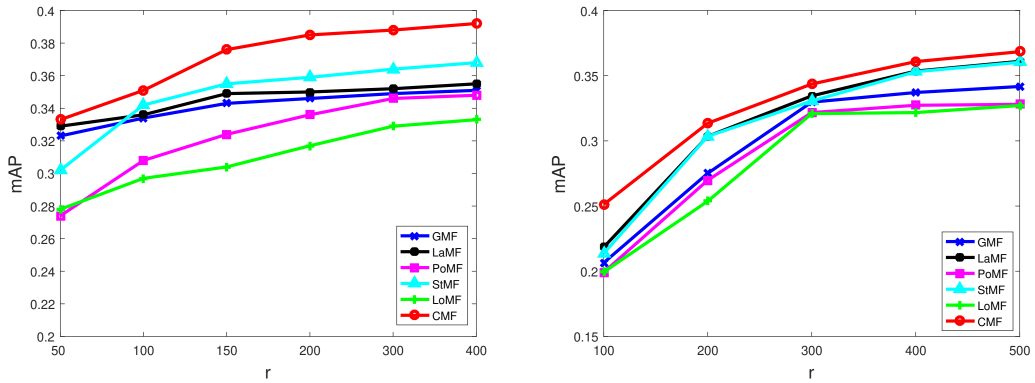
The scale parameter b in Cauchy Distribution specifies the half-width at half-maximum, which closely reflects the status of the tag noises. With the exploration over the impact of b, we could obtain the property of noises. Figure 3 demonstrates that the best bs for both datasets are around 10, which indicates the tag noises in both datasets respect to the similar distribution. When b is too small, CMF always generates a zero-matrix, where the mAP values of 0.13 in



(a) MIRFlickr

(b) NUS-WIDE

FIGURE 1: The performance of tag-based social image retrieval on the MIRFlickr and NUS-WIDE datasets in terms of mAP.



(a) MIRFlickr

(b) NUS-WIDE

FIGURE 2: The results in terms of mAP by varying the value of r on the MIRFlickr and NUS-WIDE datasets for tag-based social image retrieval.

MIRFLickr and 0.03 in NUS-WIDE are caused by the all-zero predictions. With a proper estimation of b (*i.e.* $b = 10$), CMF performs well. In addition, the too-large b also reduces the performance of CMF.

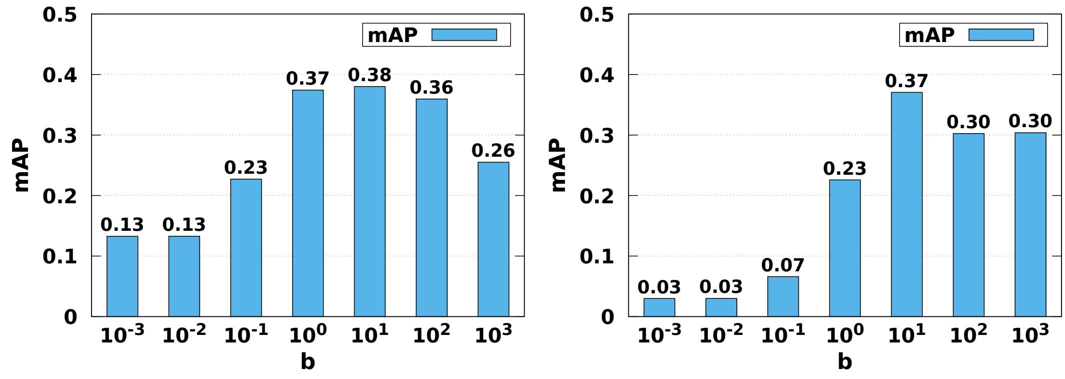
E. DISCUSSION AND EXTENSION

This work investigates the noise modeling for tag-based image retrieval, which is meaningful for many practical applications. For real-world data, it is normal that there exists noise. Thus, it is necessary to propose methods by modeling the noise. The widely-used strategy is to adopt the Gaussian noise assumption, which leads to the sum-of-squared-errors objective function. However, it is unnecessary to know the real noise probability distribution of data. Consequently, we conduct investigations about the noise assumptions and propose a new method CMF, which can well address various types of noise for tag-based image retrieval. Of course, it can be applied to many learning models. The models based on the sum-of-squared-errors objective function can be updated by introducing the new loss function, which demonstrates the important applicable and reference significance of the proposed method.

The proposed method can be easily extended. First, different regularization terms can be introduced to achieve better performance, such as the smooth regularization, the Elastic net regularization, the local structure regularization and so on. Second, the proposed method can be extended to the deep learning model by simultaneously addressing the out-of-the-sample problem as in [Li and Tang2017a], [Li et al.2019]. They enable to achieve better results for tag-based image retrieval. However, it is not our focus in this work. We focus on investigating noise modeling for tag-based image retrieval and proposing a new method based on the suitable noise assumption. How to design a better method will be researched in the future.

V. CONCLUSION

In this work, we propose a new Cauchy Matrix Factorization (CMF) method for tag-based social image retrieval. The Cauchy distribution is explored to model noise between the observed value and the ideal one. The proposed method is robust to various kinds of noise. Experiments are conducted to evaluate the effectiveness of CMF and the results demonstrate that CMF is more suitable to model the tagging noise of social images. Some extensions are also discussed. In the future, how to adaptively learn the loss function corresponding



(a) MIRFlickr

(b) NUS-WIDE

FIGURE 3: The results in terms of mAP by varying the value of b on the MIRFlickr and NUS-WIDE datasets for tag-based social image retrieval.

to the evaluation measure and adaptively set the parameters may be important research directions.

REFERENCES

[Archambeau et al.2006] C. Archambeau, N. Delannay, and M. Verleysen. Robust probabilistic projections. In ICML, 2006.

[Barnard et al.2003] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan. Matching words and pictures. JMLR, 3:1107–1135, 2003.

[Cao et al.2016] Xiangyong Cao, Qian Zhao, Deyu Meng, Yang Chen, and Zongben Xu. Robust low-rank matrix factorization under general mixture noise distributions. IEEE Transactions on Image Processing, 25(10):4677–4690, 2016.

[Chiang et al.2015] K.-Y. Chiang, C.-J. Hsieh, and I. S. Dhillon. Matrix completion with noisy side information. In NIPS, 2015.

[Chua et al.2009] T. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng. Nus-wide: A real-world web image database from national university of singapore. In ACM CIVR, 2009.

[Dong et al.2017] Yunyun Dong, Tengfei Long, Weili Jiao, Guojin He, and Zhaoming Zhang. A novel image registration method based on phase correlation using low-rank matrix factorization with mixture of gaussian. IEEE Transactions on Geoscience and Remote Sensing, 56(1):446–460, 2017.

[Du et al.2017a] Xiao-Yu Du, Yang Yang, Liu Yang, Fu-Min Shen, Zhi-Guang Qin, and Jin-Hui Tang. Captioning videos using large-scale image corpus. Journal of Computer Science and Technology, 32(3):480–493, 2017.

[Du et al.2017b] Xiaoyu Du, Jinhui Tang, Zechao Li, and Zhiguang Qin. Wheel: Accelerating cnns with distributed gpus via hybrid parallelism and alternate strategy. In Proceedings of the 25th ACM international conference on Multimedia, pages 393–401. ACM, 2017.

[Duchi et al.2011] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research, 12(Jul):2121–2159, 2011.

[Eriksson and van den Hengel2010] A. P. Eriksson and A. van den Hengel. Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L1 norm. In CVPR, 2010.

[Feng et al.2014] Z. Feng, S. Feng, R. Jin, and A. K. Jain. Image tag completion by noisy matrix recovery. In ECCV, 2014.

[Fu et al.2015] J. Fu, Y. Wu, T. Mei, J. Wang, H. Lu, and Y. Rui. Relaxing from vocabulary: Robust weakly-supervised deep learning for vocabulary-free image tagging. In ICCV, 2015.

[Gong et al.2013] Y. Gong, Q. Ke, M. Isard, and S. Lazebnik. A multi-view embedding space for modeling internet images, tags, and their semantics. IJCV, 106(2):210–233, 2013.

[Huiskes and Lew2008] M. Huiskes and M. Lew. The mir flickr retrieval evaluation. In ACM MIR, 2008.

[Kennedy et al.2006] L. S. Kennedy, S.-F. Chang, and I. Kozintsev. To search or to label?: Predicting the performance of search-based automatic image classifiers. In ACM MIR, 2006.

[Krizhevsky et al.2012] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

[Lee and Seung1999] D. Lee and H. Seung. Learning the parts of objects by nonnegative matrix factorization. Nature, 401:788–791, 1999.

[Li and Tang2017a] Z. Li and J. Tang. Weakly-supervised deep matrix factorization for social image understanding. IEEE TIP, 26(1):276–288, 2017.

[Li and Tang2017b] Z. Li and J. Tang. Weakly-supervised deep nonnegative low-rank model for social image tag refinement and assignment. In AAAI, 2017.

[Li et al.2009] X. Li, C.G.M. Snoek, and M. Worring. Learning social tag relevance by neighbor voting. IEEE TMM, 11(7):1310–1322, 2009.

[Li et al.2019] Z. Li, J. Tang, and T. Mei. Deep collaborative embedding for social image understanding. IEEE TPAMI, 41(9):2070–2083, 2019.

[Lin et al.2013] Z. Lin, G. Ding, M. Hu, J. Wang, and X. Ye. Image tag completion via image-specific and tag-specific linear sparse reconstructions. In CVPR, 2013.

[Liu et al.2010] D. Liu, X.-S. Hua, M. Wang, and H.-J. Zhang. Image retagging. In ACM Multimedia, 2010.

[Ma et al.2011] H. Ma, C. Liu, I. King, and M. R. Lyu. Probabilistic factor models for web site recommendation. In ACM SIGIR, 2011.

[Makadia et al.2010] A. Makadia, V. Pavlovic, and S. Kumar. Baselines for image annotation. IJCV, 90(1):88–105, 2010.

[Nguyen et al.2017] H. T. H. Nguyen, M. Wistuba, and L. Schmidt-Thieme. Personalized tag recommendation for images using deep transfer learning. In ECML/PKDD, 2017.

[Niu et al.2015] Y. Niu, Z. Lu, S. Huang, P. Han, and J.-R. Wen. Weakly supervised matrix factorization for noisily tagged image parsing. In IJCAI, 2015.

[Park et al.2013] S. Park, E. Serpedin, and K. A. Qaraqe. Gaussian assumption: The least favorable but the most useful. IEEE Signal Processing Magazine, 30(3):183–186, 2013.

[Salakhutdinov and Mnih2008] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In NIPS, 2008.

[Tang et al.2017] J. Tang, X. Shu, G.-J. Qi, Z. Li, M. Wang, S. Yan, and R. Jain. Tri-clustered tensor completion for social-aware image tag refinement. IEEE TPAMI, 39(8):1662–1674, 2017.

[Trigeorgis et al.2017] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller. A deep matrix factorization method for learning attribute representations. IEEE TPAMI, 39(3):417–429, 2017.

[Wong and Leung2008] R.C.F. Wong and C.H.C. Leung. Automatic semantic annotation of real-world web image. IEEE TPAMI, 30(11):1933–1944, 2008.

[Wu et al.2013] L. Wu, R. Jin, and A. K. Jain. Tag completion for image retrieval. IEEE TPAMI, 35(3):716–727, 2013.

[Xie and Xing2014] P. Xie and E. P. Xing. Cauchy principal component analysis. CoRR, abs/1412.6506, 2014.

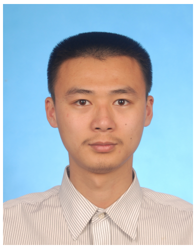
[Xue et al.2016] Z. Xue, G. Li, and Q. Huang. Joint multi-view representation learning and image tagging. In AAAI, 2016.

[Yao et al.2019] Tao Yao, Zhiwang Zhang, Lianshan Yan, Jun Yue, and Qi Tian. Discrete robust supervised hashing for cross-modal retrieval. IEEE Access, 2019.

[Zhang et al.2018] J. Zhang, Q. Wu, J. Zhang, C. Shen, and J. Lu. Kill two birds with one stone: Weakly-supervised neural network for image annotation and tag refinement. In AAAI, 2018.

[Zhu et al.2010] G. Zhu, S. Yan, and Y. Ma. Image tag refinement towards low-rank, content-tag prior and error sparsity. In ACM Multimedia, 2010.

[Znaidia et al.2013] A. Znaidia, H. Le Borgne, and C. Hudelot. Tag completion based on belief theory and neighbor voting. In ACM ICMR, 2013.



XIAOYU DU is currently a visiting scholar in the NeXT++ of National University of Singapore, and a Ph.D. candidate student of University of Electronic Science and Technology of China, Chengdu. He received his M.E. degree in computer software and theory in 2011 and B.S. degree in computer science and technology in 2008, both from Beijing Normal University, Beijing. His research interests include information retrieval, computer vision, and distributed machine learning.

QIULI LIU received the M.S. degree in Artistic Design from Nanjing Normal University at Nanjing, Jiangsu, China, in 2012. Now she is a Ph.D candidate student in University of Electronic Science and Technology of China in Software Engineering. Her research interests include multimedia theory and technology, deep learning and multimedia retrieval.

ZECHAO LI is currently a Professor at Nanjing University of Science and Technology. He received the Ph.D degree from National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences in 2013, and the B.E. degree from University of Science and Technology of China in 2008. His research interests include intelligent media analysis, computer vision, etc. He received the Young Talent Program of China Association for Science and Technology, the Excel-lent Doctoral Dissertation of Chinese Academy of Sciences, and the Excellent Doctoral Theses of China Computer Federation.

ZHIGUANG QIN is the full professor of the School of Information and Software Engineering in University of Electronic Science and Technology of China (UESTC), where he is also Director of the Key Laboratory of New Computer Application Technology and Director of UESTC-IBM Technology Center. His research interests include medical image processing, computer networking, information security, cryptography, information management, intelligent traffic, electronic commerce, distribution, and middleware.

JINHUI TANG (M'08-SM'14) received the B.Eng. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2003 and 2008, respectively. He is currently a Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. From 2008 to 2010, he was a Research Fellow with the School of Computing, National University of Singapore, Singapore. He has authored over 150 papers in top-tier journals and conferences. His current research interests include multimedia analysis and search, computer vision, and machine learning. Dr. Tang was a recipient of the best paper awards in ACM MM 2007, PCM 2011, and ICIMCS 2011, the Best Paper Runner-up in ACM MM 2015, and the best student paper awards in MMM 2016 and ICIMCS 2017. He has served as an Associate Editor for the IEEE TKDE, IEEE TNNLS, and IEEE TCSVT.

• • •