



# NRLMF $\beta$ : Beta-distribution-rescored neighborhood regularized logistic matrix factorization for improving the performance of drug–target interaction prediction



Tomohiro Ban<sup>a,b</sup>, Masahito Ohue<sup>a,c</sup>, Yutaka Akiyama<sup>a,c,d,\*</sup>

<sup>a</sup> School of Computing, Tokyo Institute of Technology, 2-12-1 W8-76 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan

<sup>b</sup> AIST-TokyoTech Real World Big-Data Computation Open Innovation Laboratory (RWBC-OIL), National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, Ibaraki, 305-8560, Japan

<sup>c</sup> Middle Molecule IT-based Drug Discovery Laboratory (MIDL), Tokyo Institute of Technology, RGBT2-A-1C 3-25-10 Tonomachi, Kawasaki-ku, Kawasaki City, Kanagawa, 210-0821, Japan

<sup>d</sup> Molecular Profiling Research Center for Drug Discovery (molprof), National Institute of Advanced Industrial Science and Technology, 2-4-7 Aomi, Koto-ku, Tokyo, 135-0064, Japan

## ARTICLE INFO

### Keywords:

Drug–target interaction prediction  
Neighborhood regularized logistic matrix factorization  
Beta distribution  
Rescoring  
Bayesian optimization  
Bayesian inference

## ABSTRACT

Techniques for predicting interactions between a drug and a target (protein) are useful for strategic drug repositioning. Neighborhood regularized logistic matrix factorization (NRLMF) is one of the state-of-the-art drug–target interaction prediction methods; it is based on a statistical model using the Bernoulli distribution. However, the prediction is not accurate when drug–target interaction pairs have less interaction information (e.g., the sum of the number of ligands for a target and the number of target proteins for a drug). This study aimed to address this issue by proposing NRLMF with beta distribution rescoring (NRLMF $\beta$ ), which is an algorithm to improve the score of NRLMF. The score of NRLMF $\beta$  is equivalent to the value of the original NRLMF score when the concentration of the beta distribution becomes infinity. The beta distribution is known as a conjugative prior distribution of the Bernoulli distribution and can reflect the amount of interaction information to its shape based on Bayesian inference. Therefore, in NRLMF $\beta$ , the beta distribution was used for rescoring the NRLMF score. In the evaluation experiment, we measured the average values of area under the receiver operating characteristics and area under precision versus recall and the 95% confidence intervals. The performance of NRLMF $\beta$  was found to be better than that of NRLMF in the four types of benchmark datasets. Thus, we concluded that NRLMF $\beta$  improved the prediction accuracy of NRLMF. The source code is available at <https://github.com/akiyamalab/NRLMFb>.

## 1. Introduction

Improving the accuracy for predicting drug–target interactions is an important task in drug discovery. Recently, research and development expenses are increasing annually, although the number of approvals for new drugs has remained constant every year [1]. Thus, increasing the value of existing drugs is necessary. Drug repositioning is one of the strategies that reuse an existing drug as a remedy for another disease [2]. For example, digoxin has been applied to prostate cancer in addition to heart failure therapy [3]. Previously, drug repositioning was casually performed during basic research and clinical trials [2]. However, it became more strategic with the development of machine learning and statistical models [4,5]. In particular, drug–target

interaction prediction is known as one of the effective approaches [6,7].

Various methods have been proposed for predicting drug–target interactions [7–15]. These methods predict unknown interactions based on known interactions and the similarity of molecular structures (e.g., Tanimoto coefficient of chemical structures and normalized Smith–Waterman score of amino acids). At present, similar drugs are assumed to interact with similar proteins and similar proteins with similar drugs. These prediction methods can be divided into two approaches. One approach involves the use of kernel method, such as WNNGIP [8], BLMNII [9], LIK [10], and ECKNN [11], and the other involves the use of a matrix factorization model, such as MSCMF [14] and NRLMF [15]. To our knowledge, NRLMF is one of the most accurate methods.

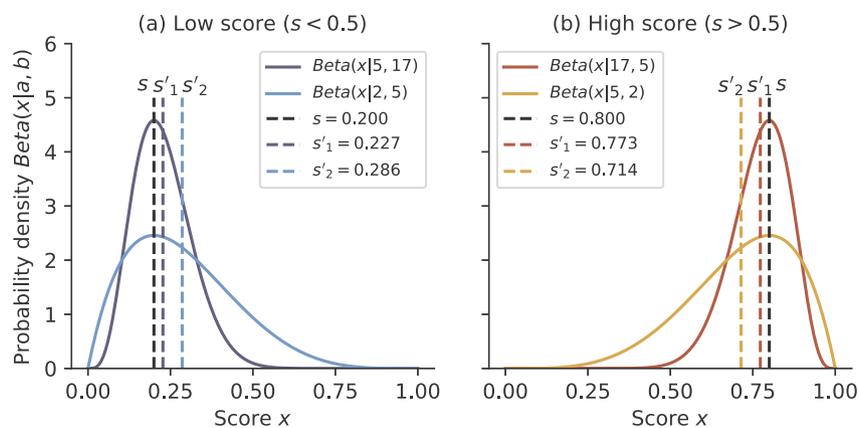
\* Corresponding author. School of Computing, Tokyo Institute of Technology, 2-12-1 W8-76 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan.  
E-mail address: [akiyama@c.titech.ac.jp](mailto:akiyama@c.titech.ac.jp) (Y. Akiyama).

<https://doi.org/10.1016/j.bbrep.2019.01.008>

Received 24 October 2018; Received in revised form 16 January 2019; Accepted 18 January 2019

Available online 07 February 2019

2405-5808/© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



**Figure 1.** A plot showing the probability density function of the beta distribution and the improved value of the score. (a) Shows the change in the score when the original score  $s$  (assuming NRLMF score) is less than 0.5.  $s'_1$  is the concentration defined by  $a + b$  (assuming the new feature  $\gamma$ ) value, and  $s'_2$  represents the score when the concentration value is small. Similarly, (b) shows the change in the score when the original score  $s$  is greater than 0.5.

However, matrix factorization models such as NRLMF cannot satisfactorily predict interactions when their number of related interactions is small. Two approaches are known to deal with this problem. One approach replaces the latent feature vectors of drugs or target proteins whose interaction number is 0 with weighted sums of latent vectors of most similar drugs or proteins [15,16]. The other adds a bias vector to each pair of drug and target protein [16]. However, with these methods, when a drug or target protein with high similarity does not exist for a drug or protein with a low number of interactions, an appropriate latent feature vector cannot be estimated. Therefore, we considered that directly correcting the score by using the number of interactions of each drug–target pair would be effective.

In this study, we proposed NRLMF with beta distribution rescaling (NRLMF $\beta$ ), which is a new drug–target interaction prediction model based on rescaling of the NRLMF score. Since NRLMF is based on the Bernoulli distribution, the beta distribution, which is its conjugate prior distribution, is used (Fig. 1). Our findings might form the basis to improve the accuracy of drug–target interaction prediction models. Herein, we introduce a new feature quantity to define the amount of interaction information and describe how to rescale the algorithm. Further, in order to confirm the improvement of prediction accuracy, we conducted a comparison experiment between the proposed method NRLMF $\beta$  and the conventional method NRLMF. In the experiment, we evaluated the area under the receiver operating characteristics (AUC) and area under precision versus recall (AUPR) by using the general benchmark [7].

## 2. Materials and methods

### 2.1. Materials

Evaluation experiments were performed using a benchmark dataset [7], which is generally used in drug–target interaction prediction [7–9,12,14,15]. The benchmark consists of four datasets targeting proteins of Nuclear Receptor, GPCR, Ion Channel, and Enzyme. Each dataset includes three matrices—an interaction matrix, a drug similarity matrix, and a protein similarity matrix. The interaction matrix is defined by an adjacency matrix consisting of 0 and 1; it takes a value of 1, if interaction is experimentally confirmed between the drug and target protein, otherwise it takes a value of 0. The interaction information was obtained from the KEGG BRITE [17], BRENDA [18], SuperTarget [19], and DrugBank [20] databases. The statistical information of the interaction matrix in each dataset is shown in Table 1. The similarity matrix of a drug is defined by a real matrix taking a value ranging from 0 to 1. The chemical structures are considered to be similar, if the value is close to 1. The structure information of drugs was

**Table 1**

Statistics for the drug–target interaction datasets.

Statistics	Nuclear receptor	GPCR	Ion channel	Enzyme
No. of drugs	54	223	210	445
No. of targets	26	95	204	664
No. of interactions	90	635	1476	2926

obtained from KEGG LIGAND [17] database. Notably, drugs having molecular weight less than 100 were excluded in order to exclude factors such as ions and cofactors. Tanimoto coefficient (i.e.,  $S_{drug}(d, d') = |d \cap d'| / |d \cup d'|$  for drugs  $d$  and  $d'$ , calculated by SIMCOMP [21]) was used for calculating the similarity between drugs. The similarity matrix of target proteins was also defined as a real matrix that takes values from 0 to 1. Amino acid sequences obtained from KEGG GENES [17] database were used for the target proteins. The similarity between target proteins was calculated using the normalized Smith–Waterman score (calculated using  $S_{target}(t, t') = SW(t, t') / \sqrt{SW(t, t)SW(t', t')}$  for targets  $t, t'$ ). Here,  $SW(\cdot, \cdot)$  represents the Smith–Waterman score [22]. These datasets can be obtained at <http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/>.

### 2.2. Problem formalization

Herein, we denote the set of drugs as  $\mathcal{D} = \{d_i\}_{i=1}^{n_d}$ , and the set of targets as  $\mathcal{T} = \{t_j\}_{j=1}^{n_t}$ , where  $n_d$  and  $n_t$  are the number of elements in sets  $\mathcal{D}$  and  $\mathcal{T}$ , respectively. The interaction matrix is represented by the adjacency matrix  $\mathbf{Y} \in \{0,1\}^{n_d \times n_t}$ . In the matrix  $\mathbf{Y}$ , the value of drug  $d_i$  and target  $t_j$  is expressed as  $y_{ij} \in \{0,1\}$ . Thus, let  $y_{ij} = 1$  (interaction pair), if interaction is observed, or let  $y_{ij} = 0$  (unknown pair). In the interaction matrix, drugs that do not show interaction with any targets are expressed as  $\mathcal{D}^- = \{d_i \in \mathcal{D} \mid \forall l, y_{il} = 0\}$  (negative drugs), whereas those that have one or more interactions are denoted as  $\mathcal{D}^+ = \mathcal{D} \setminus \mathcal{D}^-$  (positive drugs). Similarly, targets that do not have interaction with any drugs are expressed as  $\mathcal{T}^- = \{t_j \in \mathcal{T} \mid \forall l, y_{lj} = 0\}$  (negative targets), whereas those that have one or more interactions are denoted as  $\mathcal{T}^+ = \mathcal{T} \setminus \mathcal{T}^-$  (positive targets). The drug similarity matrix is expressed as  $\mathbf{S}_d$ , and the similarity between drug  $d_i$  and drug  $d_l$  is denoted as  $(S_d)_{il} \in [0,1]$ . Similarly, the target similarity matrix is expressed as  $\mathbf{S}_t$  and the similarity between target  $t_j$  and target  $t_l$  is denoted as  $(S_t)_{jl} \in [0,1]$ . This problem intends to assign high scores for drug–target pairs that have a possibility of interaction in unknown pairs.

### 2.3. NRLMF

NRLMF [15] is one of the drug–target prediction methods based on a matrix factorization technique and is one of the state-of-the-art methods.

#### 2.3.1. Interaction probability

The possibility that a drug  $d_i$  and target  $t_j$  interact is evaluated by the interaction probability  $p_{ij}$  calculated from latent feature vectors of the drug and target. The latent feature vector of drug  $d_i$  is represented by  $\mathbf{u}_i \in \mathbb{R}^r$ , and that of the target protein  $t_j$  is represented by  $\mathbf{v}_j \in \mathbb{R}^r$ , where  $r$  is a hyperparameter representing the number of dimensions of the latent feature vector.  $\mathbf{U} \in \mathbb{R}^{n_d \times r}$  is a latent matrix that has the latent feature vector  $\mathbf{u}_i^\top$  as a row vector, and  $\mathbf{V} \in \mathbb{R}^{n_t \times r}$  is a latent feature matrix that has the latent feature vector  $\mathbf{v}_j^\top$  as a row vector. Thus, the interaction probability between drug  $d_i$  and target  $t_j$  is defined as follows:

$$p_{ij} = \frac{\exp(\mathbf{u}_i^\top \mathbf{v}_j)}{1 + \exp(\mathbf{u}_i^\top \mathbf{v}_j)}. \quad (1)$$

#### 2.3.2. Prediction model

By using the interaction probability  $p_{ij}$ , we defined the likelihood of interaction matrix  $\mathbf{Y}$  in the latent feature matrix  $\mathbf{U}$ ,  $\mathbf{V}$  as follows:

$$\Pr(\mathbf{Y}|\mathbf{U}, \mathbf{V}, c) = \prod_{i=1}^{n_d} \prod_{j=1}^{n_t} p_{ij}^{cY_{ij}} (1 - p_{ij})^{1-Y_{ij}}, \quad (2)$$

where  $c > 0$  is a hyperparameter. Here, we assumed that the latent feature vector  $\mathbf{u}_i$  follows a multivariate normal distribution with mean  $0 \in \mathbb{R}^r$  and variance-covariance matrix  $\lambda_d^{-1} \mathbf{I} \in \mathbb{R}^{r \times r}$ , and then the latent feature matrix  $\mathbf{U}$  follows the joint distribution defined as equation (3). Similarly, we assumed that the latent feature vector  $\mathbf{v}_j$  follows a multivariate normal distribution with mean  $0 \in \mathbb{R}^r$  and variance-covariance matrix  $\lambda_t^{-1} \mathbf{I} \in \mathbb{R}^{r \times r}$ , and then the latent feature matrix  $\mathbf{V}$  follows the joint distribution defined as equation (4). Where  $\lambda_d, \lambda_t > 0$  are hyperparameters, and  $\mathbf{I} \in \mathbb{R}^{r \times r}$  is an identity matrix.

$$\Pr(\mathbf{U}|\lambda_d) = \prod_{i=1}^{n_d} \mathcal{N}(\mathbf{u}_i|0, \lambda_d^{-1} \mathbf{I}), \quad (3)$$

$$\Pr(\mathbf{V}|\lambda_t) = \prod_{j=1}^{n_t} \mathcal{N}(\mathbf{v}_j|0, \lambda_t^{-1} \mathbf{I}), \quad (4)$$

From equations (2)–(4), we can obtain a solution of the maximum a posteriori (MAP) by solving the following minimization problem:

$$\min_{\mathbf{U}, \mathbf{V}} \sum_{i=1}^{n_d} \sum_{j=1}^{n_t} (1 + cy_{ij} - y_{ij}) \ln[1 + \exp(\mathbf{u}_i^\top \mathbf{v}_j)] - cy_{ij} \mathbf{u}_i^\top \mathbf{v}_j + \frac{\lambda_d}{2} \|\mathbf{U}\|_F^2 + \frac{\lambda_t}{2} \|\mathbf{V}\|_F^2 \quad (5)$$

where  $\|\cdot\|_F$  is the Frobenius norm.

#### 2.3.3. Regularization

We introduced a regularization term to bring the latent feature vectors of high similarity drugs or targets close to each other by using the similarity matrices  $\mathbf{S}_d$  and  $\mathbf{S}_t$ . The regularization term in the latent feature vector of the drug is defined using the following equation.

$$\sum_{i=1}^{n_d} \sum_{l=1}^{n_t} a_{il} \|\mathbf{u}_i - \mathbf{u}_l\|_2^2 = \text{tr}(\mathbf{U}^\top \mathbf{L}_d \mathbf{U}) \quad (6)$$

$$a_{il} = \begin{cases} (\mathbf{S}_d)_{il} & \text{if } d_i \in NN_{K_1}(d_l) \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where  $\mathbf{L}_d = (\mathbf{D}^d + \tilde{\mathbf{D}}^d) - (\mathbf{A} + \mathbf{A}^\top) \in \mathbb{R}^{n_d \times n_d}$  is a real matrix.  $\mathbf{D}^d$  and  $\tilde{\mathbf{D}}^d$

are two diagonal matrices, in which the diagonal elements are  $\mathbf{D}_{ii}^d = \sum_{l=1}^{n_d} a_{il}$  and  $\tilde{\mathbf{D}}_{ii}^d = \sum_{l=1}^{n_d} a_{il}$ .  $\mathbf{A} \in \mathbb{R}^{n_d \times n_d}$  is a matrix in which the element is  $a_{il}$  in equation (7). Conversely,  $NN_{K_1}(d_i)$  is a  $K_1$  nearest neighbor decided by the  $i$ -th row vector in the matrix  $\mathbf{S}_d$ , where  $K_1 \in \mathbb{N}$  is a hyperparameter. In addition,  $\|\cdot\|_2$  is the Euclidean norm, and  $\text{tr}(\cdot)$  is the trace function. Similarly, the regularization term in the latent feature vector of the target is defined using the following equation.

$$\sum_{j=1}^{n_t} \sum_{l=1}^{n_t} b_{jl} \|\mathbf{v}_j - \mathbf{v}_l\|_2^2 = \text{tr}(\mathbf{V}^\top \mathbf{L}_t \mathbf{V}) \quad (8)$$

$$b_{jl} = \begin{cases} (\mathbf{S}_t)_{jl} & \text{if } t_l \in NN_{K_1}(t_j) \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where  $\mathbf{L}_t = (\mathbf{D}^t + \tilde{\mathbf{D}}^t) - (\mathbf{B} + \mathbf{B}^\top) \in \mathbb{R}^{n_t \times n_t}$  is a real matrix.  $\mathbf{D}^t$  and  $\tilde{\mathbf{D}}^t$  are two diagonal matrices, in which the diagonal elements are  $\mathbf{D}_{jj}^t = \sum_{l=1}^{n_t} b_{jl}$  and  $\tilde{\mathbf{D}}_{jj}^t = \sum_{l=1}^{n_t} b_{jl}$ .  $\mathbf{B} \in \mathbb{R}^{n_t \times n_t}$  is a matrix in which the element is  $b_{jl}$  in equation (9).

#### 2.3.4. Estimation and scoring

By introducing the regularization terms (6) and (8) into (5), we redefined the minimization problem as follows:

$$\min_{\mathbf{U}, \mathbf{V}} \sum_{i=1}^{n_d} \sum_{j=1}^{n_t} (1 + cy_{ij} - y_{ij}) \ln[1 + \exp(\mathbf{u}_i^\top \mathbf{v}_j)] - cy_{ij} \mathbf{u}_i^\top \mathbf{v}_j + \frac{1}{2} \text{tr}[\mathbf{U}^\top (\lambda_d \mathbf{I}_d + \alpha \mathbf{L}_d) \mathbf{U}] + \frac{1}{2} \text{tr}[\mathbf{V}^\top (\lambda_t \mathbf{I}_t + \beta \mathbf{L}_t) \mathbf{V}] \quad (10)$$

where  $\mathbf{I}_d \in \mathbb{R}^{n_d \times n_d}$  and  $\mathbf{I}_t \in \mathbb{R}^{n_t \times n_t}$  are identity matrices, and  $\alpha, \beta > 0$  are hyperparameters. Alternating gradient descent method [23] was used for optimizing equation (10). In addition, AdaGrad algorithm [24] that has a hyperparameter  $\theta > 0$  corresponding to the gradient coefficient was used to accelerate the speed of optimization. The row vectors corresponding to negative drugs and targets for the MAP solution obtained using the optimization were modified. For a negative drug  $d_i \in \mathcal{D}^-$ ,  $K_2$  nearest neighbors of the positive drugs are denoted as  $NN_{K_2}(d_i) \subset \mathcal{D}^+$ , where  $NN_{K_2}(d_i)$  is determined like  $NN_{K_1}(d_i)$  except that it targets positive drugs. Here, the latent feature vector  $\hat{\mathbf{u}}_i \in \mathbb{R}^r$  of the drug  $d_i$  corresponding to the  $i$ -th row vector of the latent feature matrix  $\hat{\mathbf{U}}$  is modified as a vector  $\tilde{\mathbf{u}}_i \in \mathbb{R}^r$  by using the following equation.

$$\tilde{\mathbf{u}}_i = \begin{cases} \hat{\mathbf{u}}_i & \text{if } d_i \in \mathcal{D}^+ \\ \frac{\sum_{d_l \in NN_{K_2}(d_i)} (\mathbf{S}_d)_{il} \hat{\mathbf{u}}_l}{\sum_{d_l \in NN_{K_2}(d_i)} (\mathbf{S}_d)_{il}} & \text{if } d_i \in \mathcal{D}^- \end{cases} \quad (11)$$

Likewise, negative target  $t_j \in \mathcal{T}^-$  is also modified. Thus, by using the modified latent feature vectors  $\tilde{\mathbf{u}}_i, \tilde{\mathbf{v}}_j \in \mathbb{R}^r$ , we calculated the score of drug  $d_i \in \mathcal{D}$  and target  $t_j \in \mathcal{T}$  by using the following equation.

$$s_{ij} = \frac{\exp(\tilde{\mathbf{u}}_i^\top \tilde{\mathbf{v}}_j)}{1 + \exp(\tilde{\mathbf{u}}_i^\top \tilde{\mathbf{v}}_j)}. \quad (12)$$

### 2.4. NRLMF $\beta$

NRLMF $\beta$  is an algorithm that rescores the score of NRLMF as the expected value of the beta distribution, which is determined based on interaction information and NRLMF score. The beta distribution is known as a conjugative prior distribution of the Bernoulli distribution used in NRLMF and can reflect the amount of interaction information to NRLMF $\beta$  score.

#### 2.4.1. New feature quantity

We defined a new feature quantity  $\gamma_{ij} \in \mathbb{Z}_+$  (where  $\mathbb{Z}_+$  is expressed as a set of integers greater than or equal to 0) as a quantity of the interaction information for any drug  $d_i$  and target  $t_j$  pair by using the

following equation.

$$\gamma_{ij} = \sum_{l=1}^{n_t} y_{il} + \sum_{l'=1}^{n_d} y_{l'j} \quad (13)$$

The feature quantity  $\gamma_{ij}$  represents the sum of the observed values of the  $i$ -th row and in the  $j$ -th column in the interaction matrix  $\mathbf{Y}$ . The  $\gamma_{ij}$  has a greater value if the number of interaction pairs in the  $i$ -th row and  $j$ -th column become larger.

#### 2.4.2. Re-scoring using the beta distribution

The NRLMF score  $s_{ij}$  based on the value of the feature quantity  $\gamma_{ij}$  becomes too small when there is little information on the interaction. To address this, we performed rescaling for the score of drug  $d_i$  and target  $t_j$  by using the beta distribution defined using the following equation.

$$\text{Beta}(x|a_{ij}, b_{ij}) = \frac{x^{a_{ij}-1}(1-x)^{b_{ij}-1}}{B(a_{ij}, b_{ij})} \quad (14)$$

where  $a_{ij}, b_{ij} > 0$  are parameters for determining the shape of the beta distribution, and the condition of  $a_{ij}, b_{ij} > 1$  was assumed in order to limit the distribution to bell shape.  $B(\cdot, \cdot)$  is a beta function. We defined the relationship between parameters  $a_{ij}, b_{ij}$ , and score  $s_{ij}$  and the feature  $\gamma_{ij}$  by using the following equation.

$$\frac{a_{ij} - 1}{a_{ij} + b_{ij} - 2} = s_{ij} \quad (15)$$

$$a_{ij} + b_{ij} = \gamma_{ij} \cdot \eta_1 + \eta_2 \quad (16)$$

where  $\eta_1, \eta_2 > 0$  are hyperparameters, and  $\eta_2 > 2$  from the above-mentioned conditions refer to the shape of the distribution. In the above equations, equation (15) refers to the mode of the beta distribution, and equation (16) refers to the concentration. Notably, the score of NRLMF corresponds to the mode of the beta distribution. From these relational expressions, the score  $s'_{ij}$  of NRLMF $\beta$  is defined as the expected value of the beta distribution by using the following equation.

$$s'_{ij} = \frac{a_{ij}}{a_{ij} + b_{ij}} = \frac{s_{ij}(\gamma_{ij} \cdot \eta_1 + \eta_2 - 2) + 1}{\gamma_{ij} \cdot \eta_1 + \eta_2} \quad (17)$$

The score  $s'_{ij}$  of NRLMF $\beta$  is equivalent to the NRLMF score  $s_{ij}$  when the concentration of the beta distribution becomes infinity. In order to prove the argument, we assume that  $\eta_1$  is at infinity and  $\eta_2$  is any finite value. Then, the argument is proved using equation (18).

$$\lim_{\eta_1 \rightarrow \infty} s'_{ij} = \lim_{\eta_1 \rightarrow \infty} \frac{s_{ij} \left( \gamma_{ij} + \frac{\eta_2}{\eta_1} - \frac{2}{\eta_1} \right) + \frac{1}{\eta_1}}{\gamma_{ij} + \frac{\eta_2}{\eta_1}} = \frac{s_{ij} \cdot \gamma_{ij}}{\gamma_{ij}} = s_{ij} \quad (18)$$

#### 2.4.3. Algorithm

The calculation procedure of NRLMF $\beta$  is shown in Algorithm 1. Initially, the observed value  $\mathbf{Y} \in \{0, 1\}^{n_d \times n_t}$  of the interaction matrix and the similarity matrices  $\mathbf{S}_d, \mathbf{S}_t$  are given as inputs. Next, the values of hyperparameters  $c, K_1, K_2, r, \lambda_d, \lambda_t, \alpha, \beta, \theta, \eta_1, \eta_2$  are set, and the NRLMF score  $s_{ij}$  is calculated for arbitrary  $i, j$  based on equation (12). For each  $i$  and  $j$ ,  $\gamma_{ij}$  is calculated from the observed value  $\mathbf{Y}$  by using equation (13). From the calculated  $\gamma_{ij}$  and NRLMF score  $s_{ij}$ , the NRLMF $\beta$  score  $s'_{ij}$  is calculated using equation (17). This algorithm rescores the score of NRLMF for all drug and protein pairs in the interaction matrix.

---

#### Algorithm 1 NRLMF $\beta$

---

Require: Observed  $\mathbf{Y} \in \{0, 1\}^{n_d \times n_t}, \mathbf{S}_d, \mathbf{S}_t$

- 1: Set hyperparameters  $c, K_1, K_2, r, \lambda_d, \lambda_t, \alpha, \beta, \theta, \eta_1, \eta_2$
  - 2: For  $\forall i, j$ , calculate NRLMF score  $s_{ij}$  by using eq. (12)
  - 3: for  $i = 1$  to  $n_d$  do
  - 4:   for  $j = 1$  to  $n_t$  do
  - 5:     Calculate  $\gamma_{ij}$  based on  $\mathbf{Y}$  by using eq. (13)
  - 6:     Calculate  $s'_{ij}$  based on  $\gamma_{ij}, s_{ij}$  by using eq. (17)
  - 7:   end for
  - 8: end for
- 

#### 2.5. Experimental settings

In order to compare the generalization performance of NRLMF and NRLMF $\beta$ , we used the AUC and AUPR as the evaluation index; according to Liu et al. [15], three kinds of 10-fold cross-validation scenarios (CVSs) called CVS1, CVS2, and CVS3 were performed five times each. The division of test data and training data is exactly the same as the comparison method NRLMF, because this evaluation result uses the script used in the original paper of the comparative method and uses the same random seed. In CVS1, all drug–target pairs  $P_1 = \{(d_i, t_j) \in \mathcal{D} \times \mathcal{T}\}$  contained in the interaction matrix  $\mathbf{Y}$  were randomly divided into ten sets so that the number of elements remained equal. A combination of one set  $P'_1 \subset P_1$  and observed values of the pair included therein were determined as the test data  $\{(d_i, t_j), y_{ij} | (d_i, t_j) \in P'_1\}$ . Further, an interaction matrix obtained by replacing all of the measured values  $y_{ij}$  of pairs included in the test data with 0 is defined as training data  $\mathbf{Y}' \in \{0, 1\}^{n_d \times n_t}$  ( $\mathbf{Y}'_{ij} = 0$  for  $\forall (d_i, t_j) \in P'_1, \mathbf{Y}'_{ij} = \mathbf{Y}_{ij}$  for  $\forall (d_i, t_j) \notin P'_1$ ). Next, each of the 10 divided sets was treated repetitively as test data, and the values of the evaluation index AUC and AUPR were measured. The above operation was repeated 5 times, and the average value of each evaluation index and the 95% confidence interval of the  $t$ -distribution were calculated. Conversely, in CVS2, the set  $\mathcal{D}$  of drugs was randomly divided into ten sets so that the number of elements was even. For a set of pairs  $P'_2 = \{(d_i, t_j) \in \mathcal{D}' \times \mathcal{T}\}$  consisting of a combination of one divided  $\mathcal{D}' \subset \mathcal{D}$  and all of the targets  $\mathcal{T}$ , a set of combinations of a pair included in the set and its observed value was determined as test data  $\{(d_i, t_j), y_{ij} | (d_i, t_j) \in P'_2\}$ . Next, like the CVS1, training data were prepared, and the value of the evaluation index was measured. Similarly, in CVS3, a set  $\mathcal{T}$  of targets was randomly divided into ten sets so that the number of elements was even. A combination of a pair included in the set and the observed value was determined as test data  $\{(d_i, t_j), y_{ij} | (d_i, t_j) \in P'_3\}$  for a pair set  $P'_3 = \{(d_i, t_j) \in \mathcal{D} \times \mathcal{T}'\}$  consisting of a combination of one divided  $\mathcal{T}'$  and all drug  $\mathcal{D}$ , and then the values of the evaluation indexes were measured.

#### 2.6. Hyperparameter optimization

In terms of the computational experiment, we cited the result of NRLMF from the original paper [15] and measured the performance of NRLMF $\beta$ . Regarding the search range of hyperparameter of NRLMF $\beta$ , based on [15], we defined the one common to NRLMF as  $c = 5, K_1 = 5, K_2 = 5, r = \{50, 100\}, \alpha = \{2^{-5}, 2^{-4}, \dots, 2^2\}, \beta,$  and  $\theta = \{2^{-5}, 2^{-4}, \dots, 2^0\}, \lambda_d = \lambda_t = \{2^{-5}, 2^{-4}, \dots, 2^1\} \{2^{-3}, 2^{-2}, \dots, 2^0\}$ . Conversely, the newly introduced hyperparameters  $\eta_1, \eta_2$  were defined as  $\eta_1 = \{2^5, 2^6, \dots, 2^9\}, \eta_2 = \{2^2, 2^3, \dots, 2^5\}$ . Regarding optimization of hyperparameters, grid search and Bayesian optimization were combined for all datasets. This is because previous experimental results [25] showed that Bayesian optimization can perform the same evaluation as a grid search. Furthermore, by dividing the search range of the hyperparameter, we can simultaneously conduct Bayesian

optimization, and the calculation time can be shortened. Therefore, the value of  $r$ ,  $\beta$ ,  $\theta$ ,  $\eta_1$ ,  $\eta_2$  was sequentially fixed based on the search range defined above, and Bayesian optimization was performed on the search range of the remaining  $\lambda_d$  and  $\alpha$  ( $\lambda_t$  is equal to  $\lambda_d$ ). Next, among solutions obtained using Bayesian optimization, those with the highest AUC average value were selected as optimal hyperparameters (Table S1). Here, we used the Gaussian process mutual information algorithm [26] as the Bayesian optimization method, and the parameter  $\delta$  was  $10^{-100}$  according to Ref. [25].

These calculations were performed using supercomputer TSUBAME 3.0 at the Tokyo Institute of Technology. Two Intel Xeon E5-2680 v4 (2.4 GHz, 14 cores) CPUs and 256 GB main memory were installed in the computing node, and calculation was performed simultaneously by using Shared Memory Parallel with seven cores (the maximum number of the `q_node` option on TSUBAME 3.0).

### 3. Results

By comparing the prediction accuracy (i.e., AUC and AUPR) of NRLMF and NRLMF $\beta$ , we showed that performance is improved by rescoring by using the beta distribution.

#### 3.1. Performance based on AUC

The average value of AUC calculated for each CVS for each dataset and the 95% confidence interval obtained using the  $t$ -distribution in order to compare the prediction accuracy of NRLMF and NRLMF $\beta$  are shown in Table 2. For CVS1 and CVS3, the average value of AUC of NRLMF $\beta$  was higher than that of NRLMF in all datasets. As for the dataset trend, both CVS1 and CVS3 were found to be remarkably improved since the dataset was smaller, indicating that NRLMF $\beta$  exerts a large effect when the amount of data is small. In particular, for the nuclear receptor dataset of CVS3, the average value of AUC of NRLMF was 0.851, whereas that of NRLMF $\beta$  was 0.941; an increase of 0.090 was observed. We will discuss this improvement in section 4.3. Conversely, for CVS2, the average value of AUC of NRLMF $\beta$  was higher than that of NRLMF for datasets other than Enzyme. However, for Enzyme, the average value of AUC of NRLMF was 0.871, whereas that of NRLMF $\beta$  was 0.858, i.e., the prediction accuracy decreased.

#### 3.2. Performance based on AUPR

The average value of AUPR calculated for each CVS for each dataset and the 95% confidence interval by the  $t$ -distribution in order to compare the prediction accuracy of NRLMF and NRLMF $\beta$  are shown in Table 3. For CVS1 and CVS3, the average value of AUPR of NRLMF $\beta$  was higher than that of NRLMF in all datasets. As for the trend of each dataset, since the dataset was smaller, the trend improved remarkably as in the case of AUC. In particular, for the nuclear receptor dataset of CVS3, the average value of AUPR of NRLMF was 0.449, whereas that of NRLMF $\beta$  was 0.661; an increase of 0.212 was observed. Conversely, for CVS2, the average value of AUPR of NRLMF $\beta$  was higher than that of NRLMF for datasets other than GPCR. For GPCR, the average value of

AUPR of NRLMF was 0.364, whereas that of NRLMF $\beta$  was 0.358, i.e., the prediction accuracy declined.

### 4. Discussion

We investigated the search range of the newly introduced hyperparameters  $\eta_1$ ,  $\eta_2$  and the reason why the performance of NRLMF $\beta$  was better than that of NRLMF.

#### 4.1. Performance based on external validation

In order to evaluate the generalization performance of the proposed method NRLMF $\beta$ , we compared it with the conventional method NRLMF by external validation (nested cross-validation). In the external validation, in order to determine the optimal hyperparameter, 5-fold cross-validation by CVS1 was performed using the training data. Next, the hyperparameter that maximizes the average AUC was determined using Bayesian optimization [25]. Subsequently, AUC and AUPR were measured using the test data by using the determined hyperparameter. In this external validation, five times of 10-fold cross-validation (CVS1) were performed using the GPCR dataset used for the cross-validation. As a result, regarding AUC, NRLMF $\beta$  was  $0.974 \pm 0.003$ , which was higher than that for NRLMF  $0.968 \pm 0.004$ . Further, for AUPR, NRLMF $\beta$  was  $0.759 \pm 0.016$ , exceeding that of NRLMF  $0.751 \pm 0.015$ . These results indicate that generalization performance can be obtained even if only amino acid sequences not including a three-dimensional structure are used. If information of the three-dimensional structures is used, improving the generalization performance would be possible.

#### 4.2. Characteristics of hyperparameters $\eta_1$ , $\eta_2$

The heatmap for the average value of AUPR when hyperparameters  $\eta_1$ ,  $\eta_2$  of NRLMF $\beta$  were changed from 2 to 512 for each pair of CVS and dataset, respectively, is shown in Fig. 2. However, for hyperparameters other than  $\eta_1$ ,  $\eta_2$ , CVS and dataset with the optimal solution were used to perform grid search with fixed  $\eta_1 = 7$  and  $\eta_2 = 3$  (see also Supplemental Information Fig. S1) by using different values for each pair of the group. The frame in each figure is the search range of the hyperparameters  $\eta_1$ ,  $\eta_2$  defined in section 2.6. This frame was created based on a heatmap that was an average of 12 heatmaps shown in Fig. 2. Based on the balance between the calculation time and prediction accuracy, we found that the total value inside the frame of size  $4 \times 5$  (where the section width of  $\eta_1$  is 5 and that of  $\eta_2$  is 4) was the largest. Thus, the frame area is  $\eta_1 = \{2^2, 2^6, \dots, 2^9\}$  and  $\eta_2 = \{2^2, 2^3, \dots, 2^5\}$ . From Fig. 2, for nuclear receptor dataset, the average value of AUPR was found to fluctuate remarkably depending on the values of  $\eta_1$ ,  $\eta_2$ . This is because the size of the dataset was small; the change in the score value for the interaction pair was thought to remarkably influence the change in the value of AUPR. Therefore, since the size of the dataset increases with GPCR, Ion channel, and Enzyme, the change in the value of AUPR with respect to the change of  $\eta_1$ ,  $\eta_2$  seemed to be small.

**Table 2**

The AUC for the 5 $\times$ 10-fold cross-validation scenarios.

Dataset	CVS1		CVS2		CVS3	
	NRLMF	NRLMF $\beta$	NRLMF	NRLMF $\beta$	NRLMF	NRLMF $\beta$
Nuclear receptor	0.950 $\pm$ 0.011	<b>0.964 <math>\pm</math> 0.007</b>	0.900 $\pm$ 0.021	<b>0.906 <math>\pm</math> 0.022</b>	0.851 $\pm$ 0.027	<b>0.941 <math>\pm</math> 0.021</b>
GPCR	0.969 $\pm$ 0.004	<b>0.975 <math>\pm</math> 0.003</b>	0.895 $\pm$ 0.011	<b>0.901 <math>\pm</math> 0.012</b>	0.930 $\pm$ 0.012	<b>0.959 <math>\pm</math> 0.006</b>
Ion channel	0.989 $\pm$ 0.001	<b>0.990 <math>\pm</math> 0.001</b>	0.813 $\pm$ 0.027	<b>0.816 <math>\pm</math> 0.027</b>	0.964 $\pm$ 0.007	<b>0.970 <math>\pm</math> 0.006</b>
Enzyme	0.987 $\pm$ 0.001	<b>0.990 <math>\pm</math> 0.001</b>	0.871 $\pm$ 0.017	0.858 $\pm$ 0.019	0.966 $\pm$ 0.005	<b>0.978 <math>\pm</math> 0.003</b>

A 10-fold cross-validation was performed five times, and the average value of AUC and the 95% confidence interval of the  $t$ -distribution are shown. The one with higher average value is shown in bold.

**Table 3**  
The AUPR for the 5×10-fold cross-validation scenarios.

Dataset	CVS1		CVS2		CVS3	
	NRLMF	NRLMF $\beta$	NRLMF	NRLMF $\beta$	NRLMF	NRLMF $\beta$
Nuclear receptor	0.728 ± 0.041	<b>0.755 ± 0.035</b>	0.545 ± 0.054	<b>0.555 ± 0.061</b>	0.449 ± 0.079	<b>0.661 ± 0.073</b>
GPCR	0.749 ± 0.015	<b>0.755 ± 0.014</b>	<b>0.364 ± 0.023</b>	0.358 ± 0.024	0.556 ± 0.038	<b>0.572 ± 0.039</b>
Ion channel	0.906 ± 0.008	<b>0.913 ± 0.008</b>	0.344 ± 0.033	<b>0.346 ± 0.033</b>	0.785 ± 0.028	<b>0.798 ± 0.027</b>
Enzyme	0.892 ± 0.006	<b>0.897 ± 0.006</b>	0.358 ± 0.040	<b>0.360 ± 0.038</b>	0.812 ± 0.018	<b>0.815 ± 0.018</b>

A 10-fold cross-validation was performed five times, and the average value of AUPR and the 95% confidence interval of the  $t$ -distribution are shown. The one with higher average value is shown in bold.

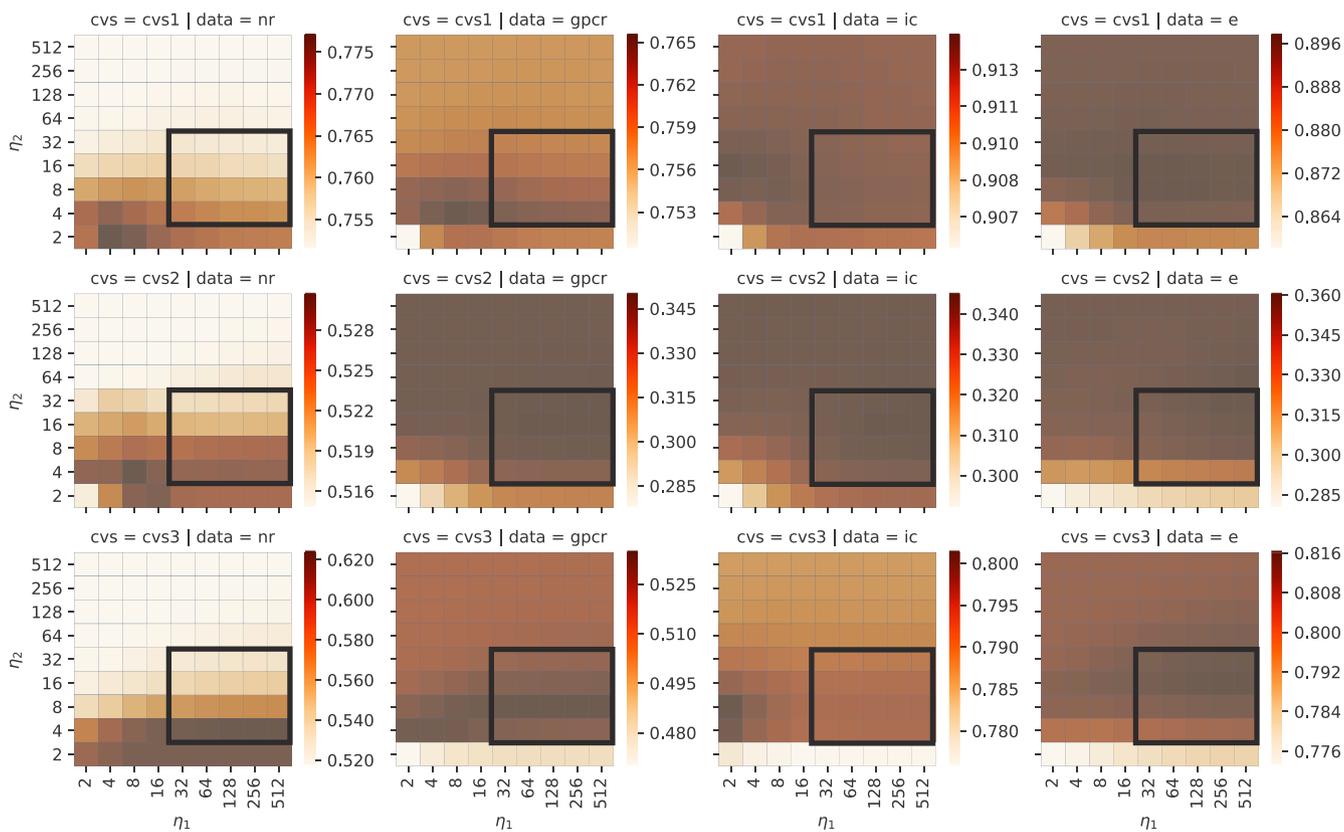
#### 4.3. Characteristics of rescaling

A plot depicting the probability density function of the beta distribution and the improved value of the score in order to explain how the score of NRLMF changes by rescaling of NRLMF $\beta$  is shown in Fig. 1. The score after improvement by rescaling of NRLMF $\beta$  differs between when the score  $s$  before improvement is low (Low score:  $s < 0.5$ ) and when it is high (High score:  $s > 0.5$ ). The change when the score before improvement is low (Low score) and the change when the concentration is different was compared (Fig. 1(a)). When the original score  $s$  was 0.200, the improved score  $s'_1$  of concentration 22 ( $a = 5, b = 17$ ) was 0.227, and the improved score  $s'_2$  of concentration 7 ( $a = 2, b = 5$ ) was 0.286. As described above, when the score  $s$  before improvement was less than 0.5, the score after improvement tended to increase; as the concentration decreases, the amount increases. Conversely, Fig. 1(b) shows the change when the score before improvement is high (High score), and the change with different concentrations is compared. When the original score  $s$  was 0.800, the improved score  $s'_1$  of concentration

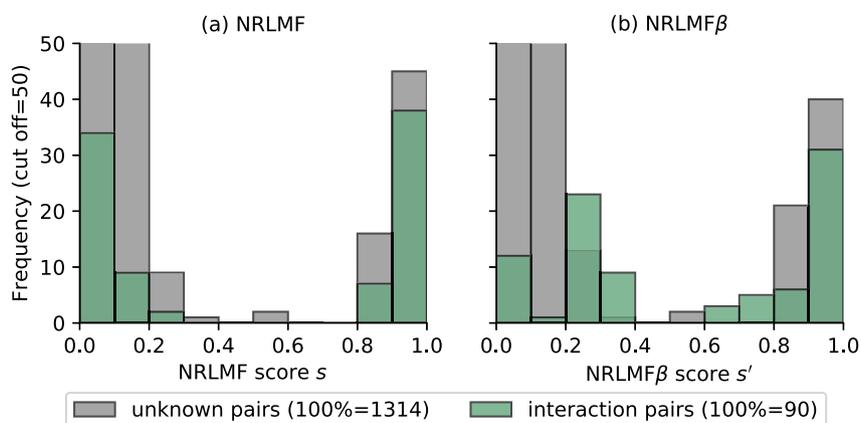
22 ( $a = 17, b = 5$ ) was 0.773, and the improved score  $s'_2$  of concentration 7 ( $a = 5, b = 2$ ) was 0.714. Thus, when the score  $s$  before improvement was larger than 0.5, the score after improvement decreased unlike that when the score before improvement was less than 0.5; when the concentration was further decreased, the amount further decreases.

#### 4.4. Discussion about improvement

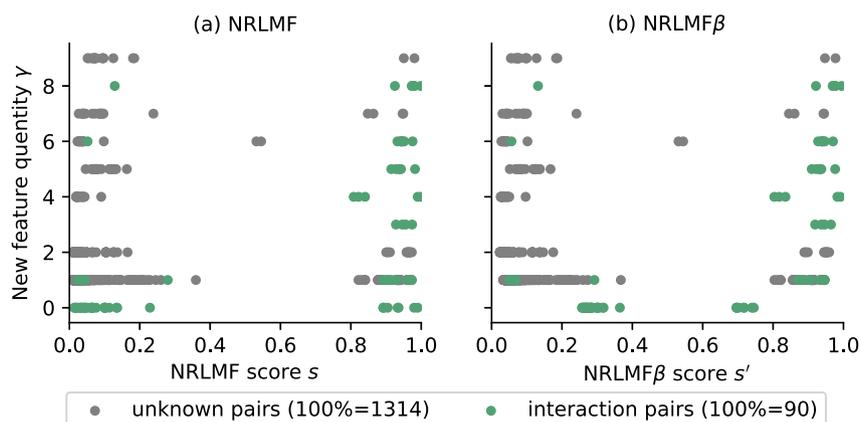
A scatter plot of scores and the feature quantity  $\gamma$  in NRLMF and NRLMF $\beta$  in order to assess the evaluation results of CVS3 for nuclear receptor dataset are shown in Fig. 3. Here, the scores  $s, s'$  and feature quantity  $\gamma$  of each method were calculated by substituting the leave-one-out method for dividing the 10-fold division of CVS3 into each target. The prediction accuracy (i.e., AUC and AUPR) of NRLMF $\beta$  improved significantly compared with that of NRLMF because the score of the interaction pairs in which the value of feature quantities  $\gamma$  is 0 was improved. In other words, the AUC and AUPR values increased as the



**Figure 2.** Heatmaps of AUPR when hyperparameters  $\eta_1, \eta_2$  are changed in each pair of cross-validation scenario (CVS) and dataset. The AUPR of each heatmap uses the average value of AUPR calculated for each CVS. However, hyperparameters without  $\eta_1, \eta_2$  were fixed using different values for each pair of CVS and dataset. In addition, the frame in each figure is the search range of the hyperparameters  $\eta_1, \eta_2$  defined in section 2.6.



**Figure 3.** Scatter plot of scores  $s$ ,  $s'$  and feature quantities  $\gamma$  of NRLMF and NRLMF $\beta$ . The score was calculated using the leave-one-out method corresponding to CVS3 for the nuclear receptor dataset. The horizontal axis represents the score of each method, and the vertical axis represents the feature quantity we introduced.



**Figure 4.** Histogram for NRLMF and NRLMF $\beta$  scores  $s$ ,  $s'$ . The score was calculated using the leave-one-out method corresponding to CVS3 for the nuclear receptor dataset. The horizontal axis represents the score of each method, and the vertical axis represents frequency (cutoff = 50).

score of interaction pairs with less than 0.5 in NRLMF was selectively increased by rescoring. This indicates that the prediction accuracy was improved by not excessively lowering the score of the pair with less information on interaction (i.e., the feature amount  $\gamma$  is small). Fig. 4 represents a histogram for each score in Fig. 3, indicating that the distribution of the score of NRLMF $\beta$  changes for NRLMF. In particular, in NRLMF, the frequency of interaction pairs with a score of 0.1 or less was found to decrease in NRLMF $\beta$ .

#### 4.5. Conclusion

Since the score of NRLMF becomes low when a pair of drugs and proteins with less interaction information is used, we proposed a resort algorithm of NRLMF $\beta$ . In NRLMF $\beta$ , the shape of the beta distribution was determined based on the value of the feature quantity  $\gamma$  representing the degree of interaction information defined by the expression (13) and the value of the score of NRLMF. The score of NRLMF was recalculated by defining the expected value of the defined beta distribution as the score of NRLMF $\beta$ . In the evaluation experiment, in order to compare the generalization performance of NRLMF and NRLMF $\beta$ , we performed three cross-validations CVS1, CVS2, and CVS3 on four datasets of Nuclear receptor, GPCR, Ion channel, and Enzyme and calculated the average values of AUC and AUPR. Hence, we confirmed that the prediction accuracy of NRLMF $\beta$  was significantly improved compared with that of NRLMF. Thus, we concluded that NRLMF $\beta$  improves the prediction accuracy for drug–target pairs with less interaction information. Future studies need to focus on the number of interactions handled within the prior distribution on the latent

feature matrix  $U, V$ . We expect that estimating more appropriate parameters and improving prediction accuracy are possible.

#### Author contributions

T.B. designed the computational method and performed the experiments. T.B. and M.O. analyzed the results. T.B., M.O., and Y.A. drafted the manuscript. Y.A. supervised the study. All authors are in agreement with the content of the manuscript for publication.

#### Competing statement

The authors declare no conflict of interest.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbrep.2019.01.008>.

#### Funding

This work was partially supported by MEXT Regional Innovation and Ecosystem Formation Program “Program to Industrialize an Innovative Middle Molecule Drug Discovery Flow through Fusion of Computational Drug Design and Chemical Synthesis Technology”; JST Research Complex Program “Wellbeing Research Campus: Creating new values through technological and social innovation”; JST CREST “Extreme Big Data” [grant number JPMJCR1303]; JSPS KAKENHI

[grant numbers 15J11261, 17H01814, and 18K18149]; and AMED BINDS [grant number JP17am0101112].

## References

- [1] Vinay Prasad, Sham Mailankody, Research and development spending to bring a single cancer drug to market and revenues after approval, *JAMA Intern. Med.* 177 (2017) 1569–1575, <https://doi.org/10.1001/jamainternmed.2017.3601>.
- [2] Ted T. Ashburn, Karl B. Thor, Drug repositioning: identifying and developing new uses for existing drugs, *Nat. Rev. Drug Discov.* 3 (2004) 673–683, <https://doi.org/10.1038/nrd1468>.
- [3] Joong Sup Shim, Jun O. Liu, Recent advances in drug repositioning for the discovery of new anticancer drugs, *Int. J. Biol. Sci.* 10 (2014) 654–663, <https://doi.org/10.7150/ijbs.9224>.
- [4] Jiao Li, Si Zheng, Bin Chen, Atul J. Butte, S. Joshua Swamidass, Zhiyong Lu, A survey of current trends in computational drug repositioning, *Briefings Bioinf.* 17 (2016) 2–12, <https://doi.org/10.1093/bib/bbv020>.
- [5] Maryam Lotfi Shahreza, Nasser Ghadiri, Sayed Rasoul Mousavi, Jaleh Varshosaz, R. James, Green, A review of network-based approaches to drug repositioning, *Briefings Bioinf.* (2017) 1–15, <https://doi.org/10.1093/bib/bbx017>.
- [6] Feixiong Cheng, Chuang Liu, Jing Jiang, Weiqiang Lu, Weihua Li, Guixia Liu, Weixing Zhou, Jin Huang, Yun Tang, Prediction of drug–target interactions and drug repositioning via network-based inference, *PLoS Comput. Biol.* 8 (2012), <https://doi.org/10.1371/journal.pcbi.1002503> e1002503.
- [7] Yoshihiro Yamanishi, Michihiro Araki, Alex Gutteridge, Wataru Honda, Minoru Kanehisa, Prediction of drug–target interaction networks from the integration of chemical and genomic spaces, *Bioinformatics* 24 (2008) i232–240, <https://doi.org/10.1093/bioinformatics/btn162>.
- [8] Twan van Laarhoven, Elena Marchiori, Predicting drug–target interactions for new drug compounds using a weighted nearest neighbor profile, *PLoS One* 8 (2013), <https://doi.org/10.1371/journal.pone.0066952> e66952.
- [9] Jian-Ping Mei, Chee-Keong Kwok, Peng Yang, Xiao-Li Li, Jie Zheng, Drug-target interaction prediction by learning from local information and neighbors, *Bioinformatics* 29 (2013) 238–245, <https://doi.org/10.1093/bioinformatics/bts670>.
- [10] Masahito Ohue, Takuro Yamazaki, Tomohiro Ban, Yutaka Akiyama, Link mining for kernel-based compound–protein interaction predictions using a chemogenomics approach, *Lect. Notes Comput. Sci.* 10362 (2017) 549–558, [https://doi.org/10.1007/978-3-319-63312-1\\_48](https://doi.org/10.1007/978-3-319-63312-1_48).
- [11] Krisztian Buza, Ladislav Peška, Drug–target interaction prediction with Bipartite Local Models and hubness-aware regression, *Neurocomputing* 260 (2017) 284–293, <https://doi.org/10.1016/j.neucom.2017.04.055>.
- [12] Mehmet Gonen, Predicting drug–target interactions from chemical and genomic kernels using Bayesian matrix factorization, *Bioinformatics* 28 (2012) 2304–2310, <https://doi.org/10.1093/bioinformatics/bts360>.
- [13] Murat Can Cobanoglu, Liu Chang, Feizhuo Hu, Zoltan N. Oltvai, Ivet Bahar, Predicting drug–target interactions using probabilistic matrix factorization, *J. Chem. Inf. Model.* 53 (2013) 3399–3409, <https://doi.org/10.1021/ci400219>.
- [14] Xiaodong Zheng, Hao Ding, Hiroshi Mamitsuka, Shanfeng Zhu, Collaborative matrix factorization with multiple similarities for predicting drug–target interactions, *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2013, pp. 1025–1033, <https://doi.org/10.1145/2487575.2487670>.
- [15] Yong Liu, Min Wu, Chunyan Miao, Peilin Zhao, Xiao-Li Li, Neighborhood regularized logistic matrix factorization for drug–target interaction prediction, *PLoS Comput. Biol.* 12 (2016), <https://doi.org/10.1371/journal.pcbi.1004760> e1004760.
- [16] Ladislav Peska, Krisztian Buza, Júlia Koller, Drug-target interaction prediction: a Bayesian ranking approach, *Comput. Methods Progr. Biomed.* 152 (2017) 15–21, <https://doi.org/10.1016/j.cmpb.2017.09.003>.
- [17] Minoru Kanehisa, Susumu Goto, Masahiro Hattori, F. Kiyoko, Aoki-Kinoshita, Masumi Itoh, Shuichi Kawashima, Toshiaki Katayama, Michihiro Araki, Mika Hirakawa, From genomics to chemical genomics: new developments in KEGG, *Nucleic Acids Res.* 34 (2006) D354–D357, <https://doi.org/10.1093/nar/gkj102>.
- [18] Schomburg Ida, Antje Chang, Christian Ebeling, Marion Gremse, Christian Heldt, Gregor Huhn, Dietmar Schomburg, BRENDA, the enzyme database: updates and major new developments, *Nucleic Acids Res.* 32 (2014) D431–D433, <https://doi.org/10.1093/nar/gkh081>.
- [19] Stefan Günther, Michael Kuhn, Mathias Dunkel, Monica Campillos, Christian Senger, Evangelia Petsalaki, Jessica Ahmed, Eduardo Garcia Urdiales, Andreas Gewiss, Lars Juhl Jensen, Reinhard Schneider, Roman Skoblo, Robert B. Russell, Philip E. Bourne, Peer Bork, Robert Preissner, SuperTarget and Matador: Resources for exploring drug–target relationships, *Nucleic Acids Res.* 36 (2008) D919–D922, <https://doi.org/10.1093/nar/gkm862>.
- [20] David S. Wishart, Craig Knox, An Chi Guo, Dean Cheng, Savita Shrivastava, Dan Tzur, Bijaya Gautam, Murtaza Hassanali, DrugBank: a knowledge base for drugs, drug actions and drug targets, *Nucleic Acids Res.* 36 (2008) D901–D906, <https://doi.org/10.1093/nar/gkm958>.
- [21] Masahiro Hattori, Yasushi Okuno, Susumu Goto, Minoru Kanehisa, Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways, *J. Am. Chem. Soc.* 125 (2003) 11853–11865, <https://doi.org/10.1021/ja036030u>.
- [22] T.F. Smith, M.S. Waterman, Identification of common molecular subsequences, *J. Mol. Biol.* 147 (1981) 195–197, [https://doi.org/10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5).
- [23] Christopher C. Johnson, Logistic matrix factorization for implicit feedback data, *Proceedings of NIPS2014 Workshop on Distributed Machine Learning and Matrix Computations*, 2014.
- [24] John Duchi, Elad Hazan, Yoram Singer, Adaptive subgradient methods for online learning and stochastic optimization, *J. Mach. Learn. Res.* 12 (2011) 2121–2159.
- [25] Tomohiro Ban, Masahito Ohue, Yutaka Akiyama, Efficient hyperparameter optimization by using Bayesian optimization for drug–target interaction prediction, *Proceedings of IEEE 7th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*, 2017, <https://doi.org/10.1109/ICCABS.2017.8114299>.
- [26] Emile Contal, Vianney Perchet, Nicolas Vayatis, Gaussian process optimization with mutual information, *Proceedings of 31st International Conference on Machine Learning (ICML2014)*, vol. 32, 2014, pp. 253–261.