BIOINFORMATICS ORIGINAL PAPER Vol. 28 no. 18 2012, pages 2304–2310
doi:10.1093/bioinformatics/bts360

Systems biology

Advance Access publication June 23, 2012

Predicting drug–target interactions from chemical and genomic kernels using Bayesian matrix factorization

Mehmet Gönen*

Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University School of Science, FI-00076 Aalto, Espoo, Finland

Associate Editor: Gunnar Ratsch

ABSTRACT

Motivation: Identifying interactions between drug compounds and target proteins has a great practical importance in the drug discovery process for known diseases. Existing databases contain very few experimentally validated drug–target interactions and formulating successful computational methods for predicting interactions remains challenging.

Results: In this study, we consider four different drug–target interaction networks from humans involving enzymes, ion channels, G-protein-coupled receptors and nuclear receptors. We then propose a novel Bayesian formulation that combines dimensionality reduction, matrix factorization and binary classification for predicting drug–target interaction networks using only chemical similarity between drug compounds and genomic similarity between target proteins. The novelty of our approach comes from the joint Bayesian formulation of projecting drug compounds and target proteins into a unified subspace using the similarities and estimating the interaction network in that subspace. We propose using a variational approximation in order to obtain an efficient inference scheme and give its detailed derivations. Finally, we demonstrate the performance of our proposed method in three different scenarios: (i) exploratory data analysis using low-dimensional projections, (ii) predicting interactions for the out-of-sample drug compounds and (iii) predicting unknown interactions of the given network.

Availability: Software and Supplementary Material are available at http://users.ics.aalto.fi/gonen/kbmf2k.

Contact: mehmet.gonen@aalto.fi

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on May 27, 2012; revised on May 27, 2012; accepted on June 18, 2012

1 INTRODUCTION

The functions of pharmaceutically useful target protein families such as enzymes, ion channels, G-protein-coupled receptors (GPCRs) and nuclear receptors can be modulated by interacting them with drug compounds. Our knowledge about the genomic space of target proteins and the chemical space of drug compounds is piling up as a result of high-throughput experimental projects that analyze the genome and high-throughput chemical compound screening with biological assays. Unfortunately, our knowledge about the relationship between these two spaces remains quite limited due to laborious and costly experimental procedures. Existing databases such as ChEMBL (Gaulton *et al.*, 2012), DrugBank (Knox *et al.*, 2011), KEGG DRUG (Kanehisa *et al.*, 2012) and SuperTarget (Hecker *et al.*, 2012) contain information about a small number of experimentally validated interactions. Hence, successful computational methods for identifying interactions between drug compounds and target proteins make genomic drug discovery significantly efficient and effective. Computational approaches can be used to guide experimentalists towards new predictions and to provide supporting evidence for their experimental results.

Traditional computational methods can be grouped into three categories: (i) docking simulations (Cheng *et al.*, 2007; Rarey *et al.*, 1996), (ii) ligand-based approaches (Butina *et al.*, 2002; Byvatov *et al.*, 2003; Keiser *et al.*, 2007) and (iii) literature text mining (Zhu *et al.*, 2005). Docking simulations require structural information of target proteins, which is not mostly available for some protein families such as GPCRs. Ligand-based approaches compare a candidate ligand with the known ligands of a target protein and may not perform well for target proteins with a small number of known ligands. Literature text mining based on keyword search cannot be used to detect unknown interactions and suffers from the redundancy due to non-standard naming practice for drug compounds and target proteins.

Recently, there are many machine learning algorithms proposed for predicting drug–target interactions using the chemical properties of drug compounds, the genomic properties of target proteins and the known interaction network (Bleakley and Yamanishi, 2009; Jacob and Vert, 2008; van Laarhoven *et al.*, 2011; Wassermann *et al.*, 2009; Yamanishi *et al.*, 2008, 2010). The main assumption of these studies is that similar drug compounds are likely to interact with similar target proteins. These similarities between drug compounds and target proteins are often encoded using kernel functions designed specifically for chemical compounds and protein sequences, respectively (Schölkopf *et al.*, 2004).

The most basic statistical approach is to formulate the interaction network inference problem as a binary classification task between drug–target pairs using pairwise kernel functions (Jacob and Vert, 2008; Wassermann *et al.*, 2009). However, this approach can be computationally quite heavy due to the high number of drug–target pairs. Supervised bipartite graph inference maps drug compounds and target points into a unified space (i.e. pharmacological space) using the chemical and genomic similarities and tries to estimate the interaction network using a distance-based procedure in that subspace (Yamanishi *et al.*, 2008, 2010). Local models are also used to predict drug–target interaction networks after their successful applications for protein–protein interaction networks, metabolic

*To whom correspondence should be addressed.

networks and regulatory networks (Bleakley and Yamanishi, 2009). Instead of using the given interaction network just as the output, integrating a kernel function that considers the given network topology and the kernels calculated using chemical compounds and protein sequences can improve the prediction performance (van Laarhoven *et al.*, 2011).

In this study, we propose a novel Bayesian formulation that combines kernel-based nonlinear dimensionality reduction (Schölkopf and Smola, 2002), matrix factorization (Srebro, 2004) and binary classification for predicting drug–target interaction networks using only chemical similarity between drug compounds and genomic similarity between target proteins. Different from previous studies, our proposed method is the first fully probabilistic formulation for drug–target interaction network inference. We show its performance on four benchmark datasets using three experimental scenarios with practical importance: (i) exploratory data analysis using low-dimensional projections, (ii) predicting interactions for the out-of-sample drug compounds and (iii) predicting unknown interactions of the given network.

2 MATERIALS

In this study, we consider four different drug–target interaction networks from humans, namely, `Enzyme`, `Ion Channel`, `GPCR` and `Nuclear Receptor`, provided by Yamanishi *et al.* (2008). These datasets are publicly available at http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/. We use these datasets as they are without adding new interactions from source databases.

2.1 Drug–target interaction data

Yamanishi *et al.* (2008) use KEGG BRITE (Kanehisa *et al.*, 2006), BRENDA (Schomburg *et al.*, 2004), SuperTarget (Günther *et al.*, 2008) and DrugBank (Wishart *et al.*, 2008) databases to collect information about the interactions between drug compounds and target proteins. Table 1 summarizes the datasets in terms of numbers of drug compounds, target proteins and interactions. The set of known drug–target interactions is regarded as 'gold standard' and used to evaluate the performance of our proposed method in the cross-validation experiments as in the previous studies (Bleakley and Yamanishi, 2009; van Laarhoven *et al.*, 2011; Yamanishi *et al.*, 2008, 2010).

2.2 Chemical data

Chemical structures of drug compounds are extracted from the DRUG and COMPOUND sections in the KEGG LIGAND database (Kanehisa *et al.*, 2006). Yamanishi *et al.* (2008) calculate the structural similarities between drug compounds using SIMCOMP (Hattori *et al.*, 2003), which represents drug compounds as graphs and calculates a similarity score based on the size of the common substructures between two graphs. Given two drug compounds $\boldsymbol{d}_i$ and $\boldsymbol{d}_k$, chemical similarity between them can be found as $s_{\mathrm{c}}(\boldsymbol{d}_i,\boldsymbol{d}_k)=|\boldsymbol{d}_i\cap\boldsymbol{d}_k|/|\boldsymbol{d}_i\cup\boldsymbol{d}_k|$ and the similarity matrix between all drug compound pairs is denoted as $\mathbf{S}_{\mathrm{c}}$.

Table 1. The drug–target interaction datasets from Yamanishi *et al.* (2008)

Dataset	Drugs	Targets	Interactions
`Enzyme`	445	664	2926
`Ion Channel`	210	204	1476
`GPCR`	223	95	635
`Nuclear Receptor`	54	26	90

2.3 Genomic data

Aminoacid sequences of target proteins are extracted from the KEGG GENES database (Kanehisa *et al.*, 2006). Yamanishi *et al.* (2008) calculate the sequence similarities between target proteins using a normalized version of Smith–Waterman score (Smith and Waterman, 1981). Given two target proteins $\boldsymbol{t}_j$ and $\boldsymbol{t}_l$, genomic similarity between them can be found as $s_{\mathrm{g}}(\boldsymbol{t}_j,\boldsymbol{t}_l)=\mathtt{SW}(\boldsymbol{t}_j,\boldsymbol{t}_l)/\sqrt{\mathtt{SW}(\boldsymbol{t}_j,\boldsymbol{t}_j)\mathtt{SW}(\boldsymbol{t}_l,\boldsymbol{t}_l)}$, where $\mathtt{SW}(\cdot,\cdot)$ gives the canonical Smith–Waterman score and the similarity matrix between all target protein pairs is denoted as $\mathbf{S}_{\mathrm{g}}$.

3 METHODS

We mainly consider the problem of predicting new drug–target interactions for out-of-sample drug compounds and/or target proteins that are not in the given interaction network. Our proposed method can also be used to predict unknown interactions of the given network.

3.1 Problem formulation

We are given $N_{\mathtt{d}}$ drug compounds denoted as $\mathbf{X}_{\mathtt{d}}=\{\boldsymbol{d}_1,\boldsymbol{d}_2,\ldots,\boldsymbol{d}_{N_{\mathtt{d}}}\}$ and $N_{\mathtt{t}}$ target proteins denoted as $\mathbf{X}_{\mathtt{t}}=\{\boldsymbol{t}_1,\boldsymbol{t}_2,\ldots,\boldsymbol{t}_{N_{\mathtt{t}}}\}$. We are also given a set of known interactions between these two sets as the $N_{\mathtt{d}}\times N_{\mathtt{t}}$ adjacency matrix denoted as $\mathbf{Y}$, where $y_j^i=+1$ if drug compound $\boldsymbol{d}_i$ interacts with target protein $\boldsymbol{t}_j$ and $y_j^i=-1$ otherwise. We can have three different prediction scenarios: (i) find interacting target proteins from $\mathbf{X}_{\mathtt{t}}$ for a new drug compound $\boldsymbol{d}_\star$, (ii) find interacting drug compounds from $\mathbf{X}_{\mathtt{d}}$ for a new target protein $\boldsymbol{t}_\star$ and (iii) estimate whether a new drug compound $\boldsymbol{d}_\star$ and a new target protein $\boldsymbol{t}_\star$ are interacting with each other or not. In order to attack these three scenarios with a single unified approach, we formulate the problem as a binary classification task, which requires to estimate whether there is an interaction between a drug compound and a target protein using only the similarities between drug compounds and the similarities between target proteins.

3.2 Kernelized Bayesian matrix factorization with twin kernels

In order to obtain an efficient Bayesian algorithm, we formulate a fully conjugate probabilistic model and develop a deterministic variational approximation mechanism for inference. The main idea is to project drug compounds and target proteins into a unified subspace using the kernels calculated from chemical and genomic data, respectively. These low-dimensional representations of drug compounds and target proteins can be used to estimate their interactions.

Figure 1 illustrates the proposed probabilistic model for predicting drug–target interactions from kernels on drug compounds and target proteins with a graphical model. The kernel matrix calculated from drug compounds $\mathbf{K}_{\mathtt{d}}$ is used to project them into a low-dimensional space using the projection matrix $\mathbf{A}_{\mathtt{d}}$. Similarly, the kernel matrix calculated from target proteins $\mathbf{K}_{\mathtt{t}}$ is used to project them into the same subspace, which is called 'pharmacological space'

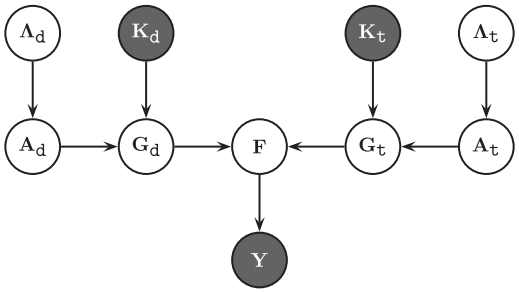

Fig. 1. Graphical model for predicting drug–target interactions from kernels on drug compounds and target proteins

in the previous studies (Yamanishi *et al.*, 2008, 2010), using the projection matrix $\mathbf{A}_{\mathtt{t}}$. The low-dimensional representations of drug compounds and target proteins in the pharmacological space, namely, $\mathbf{G}_{\mathtt{d}}$ and $\mathbf{G}_{\mathtt{t}}$, are used to calculate the interaction scores between them. Finally, the given interaction matrix $\mathbf{Y}$ is generated from the interaction score matrix $\mathbf{F}$.

The notation we use throughout the rest of the this article is as follows: $N_{\mathtt{d}}$ and $N_{\mathtt{t}}$ represent the numbers of training drug compounds and target proteins, respectively. R gives the dimensionality of the projected subspace. The $N_{\mathtt{d}} \times N_{\mathtt{d}}$ kernel matrix for drug compounds is denoted by $\mathbf{K}_{\mathtt{d}}$, where the columns of $\mathbf{K}_{\mathtt{d}}$ by $\boldsymbol{k}_{\mathtt{d},i}$. The $N_{\mathtt{d}} \times R$ matrix of corresponding projection parameters $a^i_{\mathtt{d},s}$ and their priors $\lambda^i_{\mathtt{d},s}$ are denoted by $\mathbf{A}_{\mathtt{d}}$ and $\mathbf{\Lambda}_{\mathtt{d}}$, respectively, where the columns of $\mathbf{A}_{\mathtt{d}}$ and $\mathbf{\Lambda}_{\mathtt{d}}$ by $\boldsymbol{a}_{\mathtt{d},s}$ and $\boldsymbol{\lambda}_{\mathtt{d},s}$. The $R \times N_{\mathtt{d}}$ matrix of projected instances for drug compounds $g^s_{\mathtt{d},i}$ are represented as $\mathbf{G}_{\mathtt{d}}$, where the columns of $\mathbf{G}_{\mathtt{d}}$ as $\boldsymbol{g}_{\mathtt{d},i}$ and the rows of $\mathbf{G}_{\mathtt{d}}$ as $\boldsymbol{g}^s_{\mathtt{d}}$. The $N_{\mathtt{t}} \times N_{\mathtt{t}}$ kernel matrix for target proteins is denoted by $\mathbf{K}_{\mathtt{t}}$, where the columns of $\mathbf{K}_{\mathtt{t}}$ by $\boldsymbol{k}_{\mathtt{t},j}$. The $N_{\mathtt{t}} \times R$ matrix of corresponding projection parameters $a^j_{\mathtt{t},s}$ and their priors $\lambda^j_{\mathtt{t},s}$ are denoted by $\mathbf{A}_{\mathtt{t}}$ and $\mathbf{\Lambda}_{\mathtt{t}}$, respectively, where the columns of $\mathbf{A}_{\mathtt{t}}$ and $\mathbf{\Lambda}_{\mathtt{t}}$ by $\boldsymbol{a}_{\mathtt{t},s}$ and $\boldsymbol{\lambda}_{\mathtt{t},s}$. The $R \times N_{\mathtt{t}}$ matrix of projected instances for target proteins $g^s_{\mathtt{t},j}$ are represented as $\mathbf{G}_{\mathtt{t}}$, where the columns of $\mathbf{G}_{\mathtt{t}}$ as $\boldsymbol{g}_{\mathtt{t},j}$ and the rows of $\mathbf{G}_{\mathtt{t}}$ as $\boldsymbol{g}^s_{\mathtt{t}}$. The variance for the entries of $\mathbf{G}_{\mathtt{d}}$ and $\mathbf{G}_{\mathtt{t}}$ is represented as σ_g^2. The $N_{\mathtt{d}} \times N_{\mathtt{t}}$ matrix of interaction scores f^i_j is represented as $\mathbf{F}$, where the rows of $\mathbf{F}$ as $\boldsymbol{f}^i$ and the columns of $\mathbf{F}$ as $\boldsymbol{f}_j$. The $N_{\mathtt{d}} \times N_{\mathtt{t}}$ matrix of associated interaction variables is represented as $\mathbf{Y}$, where each element $y^i_j \in \{-1,+1\}$. As short-hand notations, all priors in the model are denoted by $\mathbf{\Xi} = \{\mathbf{\Lambda}_{\mathtt{d}}, \mathbf{\Lambda}_{\mathtt{t}}\}$, where the remaining variables by $\mathbf{\Theta} = \{\mathbf{A}_{\mathtt{d}}, \mathbf{A}_{\mathtt{t}}, \mathbf{G}_{\mathtt{d}}, \mathbf{G}_{\mathtt{t}}, \mathbf{F}\}$ and the hyper-parameters by $\boldsymbol{\zeta} = \{\alpha_\lambda, \beta_\lambda\}$. Dependence on $\boldsymbol{\zeta}$ is omitted for clarity throughout this article.

The distributional assumptions of our proposed model are defined as

$$\begin{aligned}
\lambda^i_{\mathtt{d},s} &\sim \mathcal{G}(\lambda^i_{\mathtt{d},s}; \alpha_\lambda, \beta_\lambda) && \forall (i,s)\\
a^i_{\mathtt{d},s} | \lambda^i_{\mathtt{d},s} &\sim \mathcal{N}(a^i_{\mathtt{d},s}; 0, (\lambda^i_{\mathtt{d},s})^{-1}) && \forall (i,s)\\
g^s_{\mathtt{d},i} | \boldsymbol{a}_{\mathtt{d},s}, \boldsymbol{k}_{\mathtt{d},i} &\sim \mathcal{N}(g^s_{\mathtt{d},i}; \boldsymbol{a}^\top_{\mathtt{d},s} \boldsymbol{k}_{\mathtt{d},i}, \sigma_g^2) && \forall (s,i)\\
\lambda^j_{\mathtt{t},s} &\sim \mathcal{G}(\lambda^j_{\mathtt{t},s}; \alpha_\lambda, \beta_\lambda) && \forall (j,s)\\
a^j_{\mathtt{t},s} | \lambda^j_{\mathtt{t},s} &\sim \mathcal{N}(a^j_{\mathtt{t},s}; 0, (\lambda^j_{\mathtt{t},s})^{-1}) && \forall (j,s)\\
g^s_{\mathtt{t},j} | \boldsymbol{a}_{\mathtt{t},s}, \boldsymbol{k}_{\mathtt{t},j} &\sim \mathcal{N}(g^s_{\mathtt{t},j}; \boldsymbol{a}^\top_{\mathtt{t},s} \boldsymbol{k}_{\mathtt{t},j}, \sigma_g^2) && \forall (s,j)\\
f^i_j | \boldsymbol{g}_{\mathtt{d},i}, \boldsymbol{g}_{\mathtt{t},j} &\sim \mathcal{N}(f^i_j; \boldsymbol{g}^\top_{\mathtt{d},i} \boldsymbol{g}_{\mathtt{t},j}, 1) && \forall (i,j)\\
y^i_j | f^i_j &\sim \delta(f^i_j y^i_j > \nu) && \forall (i,j)
\end{aligned}$$

where the interaction scores between the matrices of projected instances and interaction variables are introduced to make the inference procedures efficient (Albert and Chib, 1993), and the margin parameter ν can be used to resolve the scaling ambiguity issue and to place a low-density region between two classes (interacting versus non-interacting), similar to the margin idea in support vector machines, which is generally used for semi-supervised learning (Lawrence and Jordan, 2005). $\mathcal{N}(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma})$ represents the normal distribution with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\mathbf{\Sigma}$. $\mathcal{G}(\cdot; \alpha, \beta)$ denotes the gamma distribution with the shape parameter α and the scale parameter β. $\delta(\cdot)$ represents the Kronecker delta function that returns 1 if its argument is true and 0 otherwise.

We only use the gamma and normal distributions in our probabilistic model. The main reason for choosing these specific distributions is that they allow us to obtain a very efficient inference mechanism easily due to conjugacy between them. Their advantage becomes clear when we explain our inference procedure.

When we consider the random variables as deterministic values, the interaction score matrix that corresponds to the decision function values in discriminative methods can be decomposed as

$$\mathbf{F} = \mathbf{G}^\top_{\mathtt{d}} \mathbf{G}_{\mathtt{t}} = (\mathbf{A}^\top_{\mathtt{d}} \mathbf{K}_{\mathtt{d}})^\top (\mathbf{A}^\top_{\mathtt{t}} \mathbf{K}_{\mathtt{t}}) = \mathbf{K}^\top_{\mathtt{d}} \mathbf{A}_{\mathtt{d}} \mathbf{A}^\top_{\mathtt{t}} \mathbf{K}_{\mathtt{t}}$$

where we see that $\mathbf{F}$ is factorized using the matrices of projected instances $\mathbf{G}_{\mathtt{d}}$ and $\mathbf{G}_{\mathtt{t}}$. Different from previous Bayesian matrix factorization solutions such as the probabilistic formulations proposed by Salakhutdinov and Mnih (2008*a*, *b*), our approach parameterizes the matrices of projected instances in terms of kernel matrices $\mathbf{K}_{\mathtt{d}}$ and $\mathbf{K}_{\mathtt{t}}$. This strategy allows us to make predictions for out-of-sample points using kernel functions.

3.2.1 Efficient inference using variational approximation Exact inference for our probabilistic model is intractable and using a Gibbs sampling approach is computationally expensive (Gelfand and Smith, 1990). We instead formulate a deterministic variational approximation, which is more efficient in terms of computation time. The variational methods use a lower bound on the marginal likelihood using an ensemble of factored posteriors to find the joint parameter distribution (Beal, 2003). We can write the factorable ensemble approximation of the required posterior as

$$p(\mathbf{\Theta}, \mathbf{\Xi} | \mathbf{K}_{\mathtt{d}}, \mathbf{K}_{\mathtt{t}}, \mathbf{Y}) \approx q(\mathbf{\Theta}, \mathbf{\Xi}) = q(\mathbf{\Lambda}_{\mathtt{d}}) q(\mathbf{A}_{\mathtt{d}}) q(\mathbf{G}_{\mathtt{d}}) q(\mathbf{\Lambda}_{\mathtt{t}}) q(\mathbf{A}_{\mathtt{t}}) q(\mathbf{G}_{\mathtt{t}}) q(\mathbf{F})$$

and define each factor just like its full conditional distribution:

$$\begin{aligned}
q(\mathbf{\Lambda}_{\mathtt{d}}) &= \prod_{i=1}^{N_{\mathtt{d}}} \prod_{s=1}^{R} \mathcal{G}(\lambda^i_{\mathtt{d},s}; \alpha(\lambda^i_{\mathtt{d},s}), \beta(\lambda^i_{\mathtt{d},s}))\\
q(\mathbf{A}_{\mathtt{d}}) &= \prod_{s=1}^{R} \mathcal{N}(\boldsymbol{a}_{\mathtt{d},s}; \mu(\boldsymbol{a}_{\mathtt{d},s}), \Sigma(\boldsymbol{a}_{\mathtt{d},s}))\\
q(\mathbf{G}_{\mathtt{d}}) &= \prod_{i=1}^{N_{\mathtt{d}}} \mathcal{N}(\boldsymbol{g}_{\mathtt{d},i}; \mu(\boldsymbol{g}_{\mathtt{d},i}), \Sigma(\boldsymbol{g}_{\mathtt{d},i}))\\
q(\mathbf{\Lambda}_{\mathtt{t}}) &= \prod_{j=1}^{N_{\mathtt{t}}} \prod_{s=1}^{R} \mathcal{G}(\lambda^j_{\mathtt{t},s}; \alpha(\lambda^j_{\mathtt{t},s}), \beta(\lambda^j_{\mathtt{t},s}))\\
q(\mathbf{A}_{\mathtt{t}}) &= \prod_{s=1}^{R} \mathcal{N}(\boldsymbol{a}_{\mathtt{t},s}; \mu(\boldsymbol{a}_{\mathtt{t},s}), \Sigma(\boldsymbol{a}_{\mathtt{t},s}))\\
q(\mathbf{G}_{\mathtt{t}}) &= \prod_{j=1}^{N_{\mathtt{t}}} \mathcal{N}(\boldsymbol{g}_{\mathtt{t},j}; \mu(\boldsymbol{g}_{\mathtt{t},j}), \Sigma(\boldsymbol{g}_{\mathtt{t},j}))\\
q(\mathbf{F}) &= \prod_{i=1}^{N_{\mathtt{d}}} \prod_{j=1}^{N_{\mathtt{t}}} \mathcal{TN}(f^i_j; \mu(f^i_j), \Sigma(f^i_j), \rho(f^i_j))
\end{aligned}$$

where $\alpha(\cdot)$, $\beta(\cdot)$, $\mu(\cdot)$ and $\Sigma(\cdot)$ denote the shape parameter, the scale parameter, the mean vector and the covariance matrix for their arguments, respectively. $\mathcal{TN}(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma}, \rho(\cdot))$ denotes the truncated normal distribution with the mean vector $\boldsymbol{\mu}$, the covariance matrix $\mathbf{\Sigma}$ and the truncation rule $\rho(\cdot)$ such that $\mathcal{TN}(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma}, \rho(\cdot)) \propto \mathcal{N}(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma})$ if $\rho(\cdot)$ is true and $\mathcal{TN}(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma}, \rho(\cdot)) = 0$ otherwise.

We can bound the marginal likelihood using Jensen's inequality:

$$\log p(\mathbf{Y} | \mathbf{K}_{\mathtt{d}}, \mathbf{K}_{\mathtt{t}}) \geq \mathrm{E}_{q(\mathbf{\Theta}, \mathbf{\Xi})}[\log p(\mathbf{Y}, \mathbf{\Theta}, \mathbf{\Xi} | \mathbf{K}_{\mathtt{d}}, \mathbf{K}_{\mathtt{t}})] - \mathrm{E}_{q(\mathbf{\Theta}, \mathbf{\Xi})}[\log q(\mathbf{\Theta}, \mathbf{\Xi})] \quad (1)$$

and optimize this bound by maximizing with respect to each factor separately until convergence. The approximate posterior distribution of a specific factor $\boldsymbol{\tau}$ can be found as

$$q(\boldsymbol{\tau}) \propto \exp(\mathrm{E}_{q(\{\mathbf{\Theta}, \mathbf{\Xi}\} \setminus \boldsymbol{\tau})}[\log p(\mathbf{Y}, \mathbf{\Theta}, \mathbf{\Xi} | \mathbf{K}_{\mathtt{d}}, \mathbf{K}_{\mathtt{t}})]).$$

For our proposed model, thanks to the conjugacy between random variables, the resulting approximate posterior distribution of each factor follows the same distribution as the corresponding factor.

3.2.2 Inference details The approximate posterior distributions of the precision priors for drug compounds can be found as

$$q(\mathbf{\Lambda}_{\mathtt{d}}) = \prod_{i=1}^{N_{\mathtt{d}}} \prod_{s=1}^{R} \mathcal{G}\left(\lambda^i_{\mathtt{d},s}; \alpha_\lambda + 1/2, (1/\beta_\lambda + \widetilde{(a^i_{\mathtt{d},s})^2}/2)^{-1}\right) \quad (2)$$

where the tilde notation denotes the posterior expectations as usual, i.e. $\widetilde{h(\boldsymbol{\tau})} = \mathrm{E}_{q(\boldsymbol{\tau})}[h(\boldsymbol{\tau})]$. The approximate posterior distribution of the projection parameters for drug compounds can be found as a product of multivariate normal distributions:

$$q(\mathbf{A}_\mathtt{d}) = \prod_{s=1}^{R} \mathcal{N}\big(\boldsymbol{a}_{\mathtt{d},s}; \Sigma(\boldsymbol{a}_{\mathtt{d},s})\mathbf{K}_\mathtt{d}\widetilde{(\boldsymbol{g}_\mathtt{d}^s)^\top}/\sigma_g^2, (\mathrm{diag}(\widetilde{\boldsymbol{\lambda}_\mathtt{d}^s}) + \mathbf{K}_\mathtt{d}\mathbf{K}_\mathtt{d}^\top/\sigma_g^2)^{-1}\big) \quad (3)$$

and the approximate posterior distribution of the projected instances for drug compounds is also a product of multivariate normal distributions:

$$q(\mathbf{G}_\mathtt{d}) = \prod_{i=1}^{N_\mathtt{d}} \mathcal{N}\big(\boldsymbol{g}_{\mathtt{d},i}; \Sigma(\boldsymbol{g}_{\mathtt{d},i})(\widetilde{\mathbf{A}_\mathtt{d}^\top}\boldsymbol{k}_{\mathtt{d},i}/\sigma_g^2 + \widetilde{\mathbf{G}_\mathtt{t}}\widetilde{(\boldsymbol{f}^i)^\top}), (\mathbf{I}/\sigma_g^2 + \widetilde{\mathbf{G}_\mathtt{t}\mathbf{G}_\mathtt{t}^\top})^{-1}\big). \quad (4)$$

The approximate posterior distributions of the precision priors for target proteins can be found as

$$q(\boldsymbol{\Lambda}_\mathtt{t}) = \prod_{j=1}^{N_\mathtt{t}}\prod_{s=1}^{R} \mathcal{G}\big(\lambda_{\mathtt{t},s}^j; \alpha_\lambda + 1/2, (1/\beta_\lambda + \widetilde{(a_{\mathtt{t},s}^j)^2}/2)^{-1}\big). \quad (5)$$

The approximate posterior distribution of the projection parameters for target proteins can be found as a product of multivariate normal distributions:

$$q(\mathbf{A}_\mathtt{t}) = \prod_{s=1}^{R} \mathcal{N}\big(\boldsymbol{a}_{\mathtt{t},s}; \Sigma(\boldsymbol{a}_{\mathtt{t},s})\mathbf{K}_\mathtt{t}\widetilde{(\boldsymbol{g}_\mathtt{t}^s)^\top}/\sigma_g^2, (\mathrm{diag}(\widetilde{\boldsymbol{\lambda}_\mathtt{t}^s}) + \mathbf{K}_\mathtt{t}\mathbf{K}_\mathtt{t}^\top/\sigma_g^2)^{-1}\big) \quad (6)$$

and the approximate posterior distribution of the projected instances for target proteins is also a product of multivariate normal distributions:

$$q(\mathbf{G}_\mathtt{t}) = \prod_{j=1}^{N_\mathtt{t}} \mathcal{N}\big(\boldsymbol{g}_{\mathtt{t},j}; \Sigma(\boldsymbol{g}_{\mathtt{t},j})(\widetilde{\mathbf{A}_\mathtt{t}^\top}\boldsymbol{k}_{\mathtt{t},j}/\sigma_g^2 + \widetilde{\mathbf{G}_\mathtt{d}}\widetilde{\boldsymbol{f}_j}), (\mathbf{I}/\sigma_g^2 + \widetilde{\mathbf{G}_\mathtt{d}\mathbf{G}_\mathtt{d}^\top})^{-1}\big). \quad (7)$$

The approximate posterior distribution of the interaction scores is a product of truncated normal distributions given as

$$q(\mathbf{F}) = \prod_{i=1}^{N_\mathtt{d}}\prod_{j=1}^{N_\mathtt{t}} \mathcal{TN}\big(f_j^i; \widetilde{\boldsymbol{g}_{\mathtt{d},i}^\top}\widetilde{\boldsymbol{g}_{\mathtt{t},j}}, 1, f_j^i y_j^i > \nu\big) \quad (8)$$

where we need to find their posterior expectations to update the approximate posterior distributions of the projected instances for drug compounds and target proteins. Fortunately, the truncated normal distribution has a closed-form formula for its expectation.

The inference procedure summarized in Algorithm 1 sequentially updates the approximate posterior distributions of the model parameters and the latent variables until convergence, which can be checked by monitoring the lower bound in (1). The first term of the lower bound corresponds to the sum of exponential form expectations of the distributions in the joint likelihood. The second term is the sum of negative entropies of the approximate posteriors in the ensemble. The only non-standard distribution in these terms is the truncated normal distribution used for the interaction scores; nevertheless, the truncated normal distribution has a closed-form formula also for its entropy.

In our implementation, the chemical similarity function $s_\mathtt{c}(\cdot,\cdot)$ is used as the kernel function between drug compounds $k_\mathtt{d}(\cdot,\cdot)$, which means that the chemical similarity matrix $\mathbf{S}_\mathtt{c}$ is used as the kernel matrix for drug compounds $\mathbf{K}_\mathtt{d}$. Similarly, the genomic similarity function $s_\mathtt{g}(\cdot,\cdot)$ is used as the kernel function between target proteins $k_\mathtt{t}(\cdot,\cdot)$, which means that the genomic similarity matrix $\mathbf{S}_\mathtt{g}$ is used as the kernel matrix for target proteins $\mathbf{K}_\mathtt{t}$. The provided similarity matrices $\mathbf{S}_\mathtt{c}$ and $\mathbf{S}_\mathtt{g}$ may not be valid kernels (i.e. positive semidefinite), but we use them as they are because our algorithm does not require them to be positive semidefinite. The hyper-parameters $(\alpha_\lambda, \beta_\lambda)$ and σ_g are set to $(1, 1)$ and 0.1, respectively.

Algorithm 1 Kernelized Bayesian matrix factorization with twin kernels (KBMF2K)

Require: $\mathbf{K}_\mathtt{d}$, $\mathbf{K}_\mathtt{t}$, $\mathbf{Y}$, R, α_λ, β_λ, σ_g and ν

1. Initialize $q(\mathbf{A}_\mathtt{d})$, $q(\mathbf{A}_\mathtt{t})$, $q(\mathbf{G}_\mathtt{d})$, $q(\mathbf{G}_\mathtt{t})$ and $q(\mathbf{F})$ randomly
2. **repeat**
3. Update $q(\boldsymbol{\Lambda}_\mathtt{d})$, $q(\mathbf{A}_\mathtt{d})$ and $q(\mathbf{G}_\mathtt{d})$ using (2), (3) and (4)
4. Update $q(\boldsymbol{\Lambda}_\mathtt{t})$, $q(\mathbf{A}_\mathtt{t})$ and $q(\mathbf{G}_\mathtt{t})$ using (5), (6) and (7)
5. Update $q(\mathbf{F})$ using (8)
6. **until** convergence
7. **return** $q(\mathbf{A}_\mathtt{d})$ and $q(\mathbf{A}_\mathtt{t})$

3.3 Prediction scenarios

We consider three different scenarios for drug–target interaction prediction. For these scenarios, we can get probabilistic estimates from our Bayesian model but the variances are observed to be very small due to discriminative nature of the model (i.e. modeling the interaction between drug compounds and target proteins by introducing the binary classification part with a large margin strategy just after the matrix factorization part). Hence, we only consider point estimates for simplicity without sacrificing the generalization performance.

3.3.1 Prediction for a new drug compound In the first scenario, we assume that we are given a new drug compound $\boldsymbol{d}_\star$ and our task is to find the set of target proteins from $\mathbf{X}_\mathtt{t}$ that interact with $\boldsymbol{d}_\star$. We first need to calculate the similarities between $\boldsymbol{d}_\star$ and $\mathbf{X}_\mathtt{d}$:

$$\boldsymbol{k}_{\mathtt{d},\star} = \big[k_\mathtt{d}(\boldsymbol{d}_\star, \boldsymbol{d}_1)\ k_\mathtt{d}(\boldsymbol{d}_\star, \boldsymbol{d}_2)\ \ldots\ k_\mathtt{d}(\boldsymbol{d}_\star, \boldsymbol{d}_{N_\mathtt{d}})\big]^\top$$

and these similarities can be used to find the interaction scores for $\boldsymbol{d}_\star$:

$$\boldsymbol{f}^\star = \big(\widetilde{\mathbf{A}_\mathtt{d}^\top}\boldsymbol{k}_{\mathtt{d},\star}\big)^\top\big(\widetilde{\mathbf{A}_\mathtt{t}^\top}\mathbf{K}_\mathtt{t}\big) = \boldsymbol{k}_{\mathtt{d},\star}^\top\widetilde{\mathbf{A}_\mathtt{d}}\widetilde{\mathbf{A}_\mathtt{t}^\top}\mathbf{K}_\mathtt{t}$$

where positive valued entries indicate that the corresponding target proteins interact with $\boldsymbol{d}_\star$.

3.3.2 Prediction for a new target protein In the second scenario, we assume that we are given a new target protein $\boldsymbol{t}_*$ and our task is to find the set of drug compounds from $\mathbf{X}_\mathtt{d}$ that interact with $\boldsymbol{t}_*$. We first need to calculate the similarities between $\boldsymbol{t}_*$ and $\mathbf{X}_\mathtt{t}$:

$$\boldsymbol{k}_{\mathtt{t},*} = \big[k_\mathtt{t}(\boldsymbol{t}_*, \boldsymbol{t}_1)\ k_\mathtt{t}(\boldsymbol{t}_*, \boldsymbol{t}_2)\ \ldots\ k_\mathtt{t}(\boldsymbol{t}_*, \boldsymbol{t}_{N_\mathtt{t}})\big]^\top$$

and these similarities can be used to find the interaction scores for $\boldsymbol{t}_*$:

$$\boldsymbol{f}_* = \big(\widetilde{\mathbf{A}_\mathtt{d}^\top}\mathbf{K}_\mathtt{d}\big)^\top\big(\widetilde{\mathbf{A}_\mathtt{t}^\top}\boldsymbol{k}_{\mathtt{t},*}\big) = \mathbf{K}_\mathtt{d}^\top\widetilde{\mathbf{A}_\mathtt{d}}\widetilde{\mathbf{A}_\mathtt{t}^\top}\boldsymbol{k}_{\mathtt{t},*}$$

where positive valued entries indicate that the corresponding drug compounds interact with $\boldsymbol{t}_*$.

3.3.3 Joint prediction for a new drug compound and a new target protein The third scenario is the hybrid of the first two scenarios and our task is to find whether a new drug compound $\boldsymbol{d}_\star$ and a new target protein $\boldsymbol{t}_*$ interact with each other. We can find the interaction score for $\boldsymbol{d}_\star$ and $\boldsymbol{t}_*$ as

$$f_*^\star = \big(\widetilde{\mathbf{A}_\mathtt{d}^\top}\boldsymbol{k}_{\mathtt{d},\star}\big)^\top\big(\widetilde{\mathbf{A}_\mathtt{t}^\top}\boldsymbol{k}_{\mathtt{t},*}\big) = \boldsymbol{k}_{\mathtt{d},\star}^\top\widetilde{\mathbf{A}_\mathtt{d}}\widetilde{\mathbf{A}_\mathtt{t}^\top}\boldsymbol{k}_{\mathtt{t},*}$$

where a positive value indicates that $\boldsymbol{d}_\star$ and $\boldsymbol{t}_*$ interact with each other.

Note that $\boldsymbol{d}_\star$ and $\boldsymbol{t}_*$ can be a known drug compound and a known target protein, respectively, from the given interaction network in order to predict unknown interactions.

4 RESULTS

In order to illustrate the effectiveness of our proposed method, called 'kernelized Bayesian matrix factorization with twin kernels' (KBMF2K), we present the results of three experimental scenarios: (i) exploratory data analysis using low-dimensional projections, (ii) predicting interactions for the out-of-sample drug compounds and (iii) predicting unknown interactions of the given network.

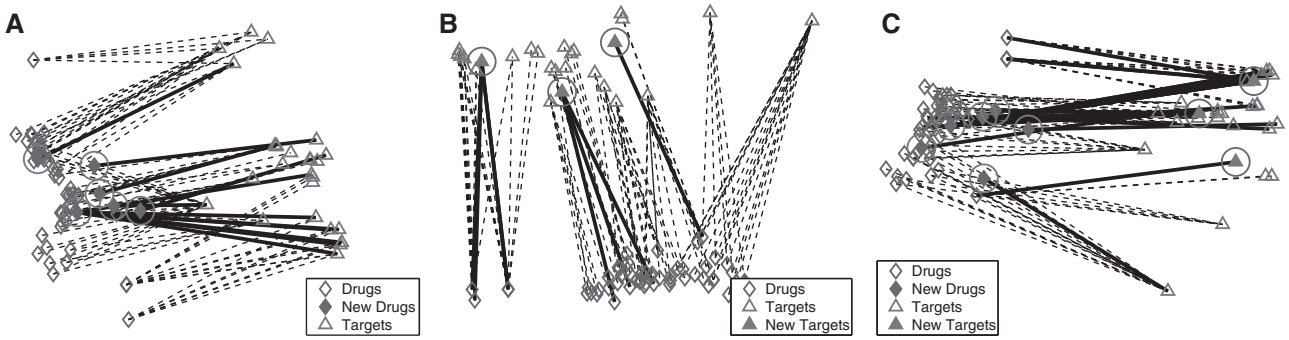

Fig. 2. Two-dimensional projections of drug compounds and target proteins obtained by KBMF2K on Nuclear Receptor dataset with (**A**) held-out drug compounds (**B**) held-out target proteins compounds and (**C**) held-out drug compounds and target proteins. Provided interactions between drug compounds and target proteins are shown as dashed lines for training network and thick solid lines for held-out drug compounds and/or target proteins

4.1 Exploratory data analysis using low-dimensional projections

KBMF2K can also be used for exploratory data analysis by displaying low-dimensional projections in addition to predicting interactions. For three different prediction scenarios described earlier, we provide visualizations on Nuclear Receptor dataset due to its small network size. In this set of experiments, we set the subspace dimensionality $R=2$ and the margin parameter $\nu=0$.

Given a drug–target interaction network, we want to investigate the interactions of new drug compounds and/or target proteins within that network. We do not include 10% of drug compounds and/or target proteins and their interactions to our training network giving us three different scenarios. Figure 2 displays the two-dimensional projections of training networks superimposed with the predicted projections for held-out drug compounds and/or target proteins. Provided interactions between drug compounds and target proteins are also shown as dashed lines for training network and thick solid lines for held-out drug compounds and/or target proteins.

We would like to point out a couple of important observations from Figure 2. First of all, KBMF2K successfully captures bipartite nature of the given interaction networks (i.e. two disjoint node sets) by placing drug compounds and target proteins as clearly separated node groups. Second, we can easily see that the dashed lines (i.e. interactions from training network) connect nearby drug compounds and target proteins. Finally the projections for held-out drug compounds/target proteins are meaningful because they are connected to nearby target proteins/drug compounds.

The prediction performance using just two dimensions may not be enough for a reliable system, but these two-dimensional figures can definitely be used for exploratory data analysis.

4.2 Predicting interactions for the out-of-sample drug compounds

To show the performance of KBMF2K in predicting interactions for new drug compounds, we perform experiments on the four benchmark datasets. We exactly follow the experimental procedure of Yamanishi *et al.* (2010) in order to have comparable results. For each dataset, drug compounds are split into five subsets of roughly equal size. Each subset is then used in turn as the test set and training is performed on the remaining four subsets. This procedure is repeated five times to obtain robust results. The subspace dimensionality R and the margin parameter ν of KBMF2K are selected from $\{5,10,15,20,25\}$ and $\{0,1\}$, respectively, using the prediction performances on the training sets.

Table 2 gives the average AUC (area under the receiver operating curve) values for Yamanishi *et al.* (2010) and KBMF2K. Note that Yamanishi *et al.* (2010) also report results with pharmacological similarity between drug compounds. We compare our results with the results obtained using the same similarity measures in our experiments (i.e. chemical similarity for drug compounds and genomic similarity for target proteins). We see that KBMF2K achieves higher average AUC values on all datasets. KBMF2K significantly improves the results on Ion Channel and GPCR datasets by 10.7% and 4.6% respectively.

Figure 3 shows the average AUC values for KBMF2K with changing subspace dimensionality and $\nu=0$. On Nuclear Receptor dataset, we do not see any effect of the subspace dimensionality possibly due to small size of the interaction network. However, there is a clear increasing trend in AUC with increasing subspace dimensionality for other datasets. On Enzyme and GPCR datasets, we get the best results with $R=25$. It is still possible to improve the results on Enzyme dataset by adding more dimensions to the common subspace of drug compounds and target proteins.

Instead of using a cross-validation strategy, the intrinsic subspace dimensionality can be found while learning the model parameters using, for example, automatic relevance determination (Neal, 1996). However, we leave this extension as future work.

Table 2. Prediction performances of Yamanishi *et al.* (2010) and KBMF2K on the four benchmark datasets in terms of average AUC values

Dataset	Yamanishi *et al.* (2010)	KBMF2K
Enzyme	0.821	0.832
Ion Channel	0.692	0.799
GPCR	0.811	0.857
Nuclear Receptor	0.814	0.824

Table 3. The top five predicted interactions on the four benchmark datasets

	Rank	Pair	Annotation		Rank	Pair	Annotation
Enzyme	**1**	**D00437**	**Nifedipine (JP16/USP/INN)**	GPCR	**1**	**D02358**	**Metoprolol (USAN/INN)**
	CD	**1559**	**cytochrome P450, family 2, subfamily C, polypeptide 9**		**CD**	**154**	**adrenergic, beta-2-, receptor, surface**
	2	**D00542**	**Halothane (JP16/USP/INN)**		**2**	**D04625**	**Isoetharine (USP)**
	CDK	**1571**	**cytochrome P450, family 2, subfamily E, polypeptide 1**		**C K**	**154**	**adrenergic, beta-2-, receptor, surface**
	3	**D00097**	**Salicylic acid (JP16/USP)**		**3**	**D00283**	**Clozapine (JAN/USP/INN)**
	CD	**5743**	**prostaglandin-endoperoxide synthase 2**		**CD**	**1814**	**dopamine receptor D3**
	4	D00501	Pentoxifylline (JAN/USP/INN)		4	D02354	Thiethylperazine (USAN/INN)
		5150	phosphodiesterase 7A			1814	dopamine receptor D3
	5	**D00139**	**Methoxsalen (JP16/USP)**		**5**	**D00604**	**Clonidine hydrochloride (JP16/USP)**
	DK	**1543**	**cytochrome P450, family 1, subfamily A, polypeptide 1**		**C**	**148**	**adrenergic, alpha-1A-, receptor**
Ion Channel	**1**	**D00438**	**Nimodipine (USAN/INN)**	Nuclear Receptor	**1**	**D00182**	**Norethisterone (JP16)**
	DK	**779**	**calcium channel, voltage-dependent, L type, alpha 1S subunit**		**C**	**2099**	**estrogen receptor 1**
	2	**D00538**	**Zonisamide (JAN/USAN/INN)**		**2**	**D00348**	**Isotretinoin (USP)**
	DK	**6331**	**sodium channel, voltage-gated, type V, alpha subunit**		**C**	**5915**	**retinoic acid receptor, beta**
	3	**D00552**	**Ethyl aminobenzoate (JP16)**		**3**	**D00348**	**Isotretinoin (USP)**
	K	**6331**	**sodium channel, voltage-gated, type V, alpha subunit**		**C**	**5916**	**retinoic acid receptor, gamma**
	4	D00546	Desflurane (JAN/USP/INN)		**4**	**D00898**	**Dienestrol (USP/INN)**
		2555	gamma-aminobutyric acid (GABA) A receptor, alpha 2		**C K**	**2100**	**estrogen receptor 2**
	5	D00528	Anhydrous caffeine (JP16)		**5**	**D00348**	**Isotretinoin (USP)**
		1080	cystic fibrosis transmembrane conductance regulator		**C**	**6258**	**retinoid X receptor, gamma**

Interactions reported in ChEMBL, DrugBank and KEGG are marked with C, D and K, respectively. Interactions reported in at least one of these databases are shown in bold.

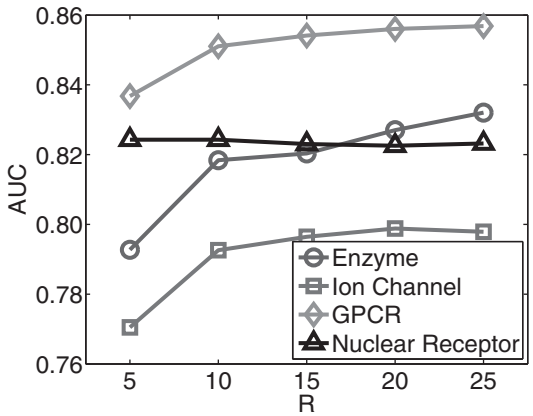

Fig. 3. Prediction performance of KBMF2K with changing subspace dimensionality and $\nu = 0$ on the four benchmark datasets in terms of average AUC values

4.3 Predicting unknown interactions of the given network

In order to illustrate the performance of KBMF2K in predicting unknown drug–target interactions of the given network, we perform a new set of experiments on the four benchmark datasets. Using the best parameter values for $\{R, \nu\}$ found in the previous experiments, we train KBMF2K with the complete interaction network for each dataset. We rank the non-interacting pairs with respect to their interaction scores and extract the top 100 predicted interactions. We report only the top five predicted interactions for each dataset and give the full lists of predicted interactions as Supplementary material.

Table 3 lists the top five predicted interactions for each dataset. We check these predicted interactions manually from the latest online versions of ChEMBL (Gaulton *et al.*, 2012), DrugBank (Knox *et al.*, 2011) and KEGG DRUG (Kanehisa *et al.*, 2012) databases. We see that 80% of the predictions (16 out of 20) is reported in at least one of these databases. This is a strong evidence for the practical relevance of our method. For example, Enzyme dataset has 2926 interacting and 292554 non-interacting (i.e. not known to interact) drug–target pairs. We pick only the top five predicted interactions and see that four out of these five drug–target pairs are currently reported in at least one database. KBMF2K correctly identifies three and four out of five predicted interactions on Ion Channel and GPCR datasets, respectively. The prediction performance of KBMF2K is even better on Nuclear Receptor dataset. The top five predicted interactions are currently reported in ChEMBL database. We also check the top 10 predicted interactions and see that all of them are reported in ChEMBL database. Note that the predicted interactions that are not reported yet may also exist in reality.

5 DISCUSSION

In this study, we consider four different drug–target interaction networks from humans involving enzymes, ion channels, GPCRs and nuclear receptors. We then propose a novel Bayesian formulation that combines kernel-based nonlinear dimensionality reduction (Schölkopf and Smola, 2002), matrix factorization (Srebro, 2004) and binary classification for predicting drug–target interaction networks using only chemical similarity between drug compounds and genomic similarity between target proteins. The novelty of our approach comes from the joint Bayesian formulation of projecting drug compounds and target proteins into a unified subspace using the similarities and estimating the interaction

network in that subspace. Our proposed method is the first fully probabilistic formulation proposed for drug–target interaction network inference.

We propose using a variational approximation in order to obtain an efficient inference scheme and give its detailed derivations. The most time-consuming steps of the proposed variational inference mechanism are covariance calculations because we need to perform matrix inversions. The time complexity of the covariance updates for the projection matrices in (3) and (6) is $\mathcal{O}(RN_{\mathtt{d}}^3)$ and $\mathcal{O}(RN_{\mathtt{t}}^3)$, respectively. The time complexity of the covariance updates for the composite components in (4) and (7) is $\mathcal{O}(R^3)$. The other calculations in these steps can be done very efficiently using matrix–matrix or matrix–vector multiplications. Finding the posterior expectations of the interaction scores in (8) only requires evaluating the standardized normal cumulative distribution function and the standardized normal probability density. In summary, the total time complexity of each iteration in our variational approximation scheme is $\mathcal{O}(RN_{\mathtt{d}}^3+RN_{\mathtt{t}}^3+R^3)$, which makes our algorithm very efficient compared to standard pairwise kernel approaches that require calculating an $N_{\mathtt{d}}N_{\mathtt{t}} \times N_{\mathtt{d}}N_{\mathtt{t}}$ kernel matrix between object pairs and training a kernel-based classifier using this kernel matrix.

In order to demonstrate the performance of our proposed method, called 'kernelized Bayesian matrix factorization with twin kernels' (`KBMF2K`), we use four benchmark datasets containing known drug–target interaction networks, chemical kernels between drug compounds and genomic kernels between target proteins provided by Yamanishi *et al.* (2008). We design three different experimental scenarios with practical importance as follows (i) exploratory data analysis using low-dimensional projections, (ii) predicting interactions for the out-of-sample drug compounds and (iii) predicting unknown interactions of the given network. In the first set of results, we show that the resulting low-dimensional projections can be used to predict drug–target interactions and practitioners can use these projections as two-dimensional figures for exploratory data analysis. The remaining sets of results show that our novel probabilistic interpretation obtains better generalization performance than earlier optimization-based approaches.

`KBMF2K` uses one kernel function for chemical similarity and another kernel function calculated on protein sequences for genomic similarity. The performance of our approach can be improved by integrating multiple kernels for both kinds of similarity. In kernel-based methods, this approach is known as 'multiple kernel learning' (Gönen and Alpaydın, 2011) and our method can be extended towards that direction.

ACKNOWLEDGEMENT

The author thanks Fidan Sümbül for her very useful comments and suggestions.

Funding: The Academy of Finland (Finnish Centre of Excellence in Computational Inference Research COIN, grant no 251170).

Conflict of interest: none declared.

REFERENCES

Albert,J.H. and Chib,S. (1993) Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.,* **88**, 669–679.

Beal,M.J. (2003) *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, The Gatsby Computational Neuroscience Unit, University College London.

Bleakley,K. and Yamanishi,Y. (2009) Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics,* **25**, 2397–2403.

Butina,D. *et al.* (2002) Predicting ADME properties *in silico*: methods and models. *Drug Discov. Today,* **7**, S83–S88.

Byvatov,E. *et al.* (2003) Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J. Chem. Inf. Comput. Sci.,* **43**, 1882–1889.

Cheng,A.C. *et al.* (2007) Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.*, **25**, 71–75.

Gaulton,A. *et al.* (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.,* **40**, D1100–D1107.

Gelfand,A.E. and Smith,A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.,* **85**, 398–409.

Gönen,M. and Alpaydın,E. (2008) Multiple kernel learning algorithms. *J. Mach. Learn. Res.,* **12**, 2211–2268.

Günther,S. *et al.* (2008) SuperTarget and Matador: resources for exploring drug–target relationships. *Nucleic Acids Res.,* **36**, D919–D922.

Hattori,M. *et al.* (2003) Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.,* **125**, 11853–11865.

Hecker,N. *et al.* (2012) SuperTarget goes quantitative: update on drug–target interactions. *Nucleic Acids Res.,* **40**, D1113–D1117.

Jacob,L. and Vert,J.-P. (2008) Protein–ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics,* **24**, 2149–2156.

Kanehisa,M. *et al.* (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.,* **34**, D354–D357.

Kanehisa,M. *et al.* (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.,* **40**, D109–D114.

Keiser,M.J. *et al.* (2007) Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.,* **25**, 197–206.

Knox,C. *et al.* (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.,* **39**, D1035–D1041.

Lawrence,N.D. and Jordan,M.I. (2005) Semi-supervised learning via Gaussian processes. In *Advances in Neural Information Processing Systems 17*, pp. 753–760.

Neal,R.M. (1996) *Bayesian Learning for Neural Networks*. Springer, New York, NY.

Rarey,M. *et al.* (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.,* **261**, 470–489.

Salakhutdinov,R. and Mnih,A. (2008*a*) Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *Proceedings of the 25th International Conference on Machine Learning*, pp. 880–887.

Salakhutdinov,R. and Mnih,A. (2008*b*) Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems 20*, pp. 1257–1264.

Schölkopf,B. and Smola,A.J. (2002) *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA.

Schölkopf,B. *et al.*, (eds) (2004) *Kernel Methods in Computational Biology*. MIT Press, Cambridge, MA.

Schomburg,I. *et al.* (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.,* **32**, D431–D433.

Smith,T.F. and Waterman,M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.,* **147**, 195–197.

Srebro,N. (2004) *Learning with Matrix Factorizations*. PhD thesis, Massachusetts Institute of Technology.

van Laarhoven,T. *et al.* (2011) Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics,* **27**, 3036–3043.

Wassermann,A.M. *et al.* (2009) Ligand prediction for orphan targets using support vector machines and various target–ligand kernels is dominated by nearest neighbor effects. *J. Chem. Inf. Model.,* **49**, 2155–2167.

Wishart,D.S. *et al.* (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.,* **36**, D901–D906.

Yamanishi,Y. *et al.* (2008) Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics,* **24**, i232–i240.

Yamanishi,Y. *et al.* (2010) Drug–target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics,* **26**, i246–i254.

Zhu,S. *et al.* (2005) A probabilistic model for mining implicit 'chemical compound-gene' relations from literature. *Bioinformatics,* **21** (Suppl 2), ii245–ii251.