Knowledge-Based Systems 199 (2020) 105978

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

MTMA: Multi-task multi-attribute learning for the prediction of adverse drug–drug interaction

Jiajing Zhu [a], Yongguo Liu [a,*], Chuanbiao Wen [b]

[a] Knowledge and Data Engineering Laboratory of Chinese Medicine, School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China
[b] College of Medical Information Engineering, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

ARTICLE INFO

Article history:
Received 27 October 2019
Received in revised form 24 April 2020
Accepted 26 April 2020
Available online 30 April 2020

Keywords:
Adverse drug–drug interaction
Multi-task
Multi-attribute
Supervised learning
Tensor decomposition

ABSTRACT

Adverse drug–drug interaction (ADDI) is an important issue in drug developments and clinical applications, which causes a significant burden in the healthcare system and leads to serious morbidity and mortality in patients. Many methods are proposed for ADDI prediction due to the accumulation of healthcare data in a massive scale. However, these methods are insufficient in exploring the potential adverse mechanisms among drugs and incapable of revealing the leading factors of ADDIs. In this paper, we propose a Multi-Task Multi-Attribute (MTMA) learning model for ADDI prediction. In MTMA, two drug attributes, molecular structure and side effect, are adopted to model the adverse interactions among drugs and two interpretable tensors, adverse molecular structure–molecular structure interaction tensor and adverse side effect-side effect interaction tensor, are designed to uncover the adverse mechanisms among drugs. Meanwhile, we impose $l_{2,1}$-norm on the predicted attribute matrices to explore the leading molecular substructures and side effects for each specific ADDI. The optimization problem of MTMA is solved by an alternating algorithm based on the methods of low-rank tensor decomposition and stochastic gradient descent. Experiments on the real-world dataset demonstrate the considerable performance of MTMA when compared with nine baseline methods and its three variants.

1. Introduction

Adverse drug reaction (ADR) is known as unpleasant or harmful effects caused by individual drugs and drug combinations, which has become a significant threat to public health [1–3]. ADR is estimated to result in more than 2 million hospitalizations and 100,000 deaths per year and wipe about $136 billion from healthcare annual budget in the United States [4,5]. As one of the leading causes of ADR, adverse drug–drug interaction (ADDI) is referred to as the adverse reactions or dangerous effects caused by the co-administration of two drugs [6,7], which is responsible for 30% of the reported ADR events and becomes an important clinical problem and a significant financial burden for healthcare system [8–10]. For example, *Aspirin*, an analgesic–antipyretic, taken together with *Warfarin*, a blood-thinning drug, may result in gastrointestinal hemorrhage, excessive bleeding, and urethrorrhagia [3]. Although the detection of ADDIs is conducted during the preclinical stage of drug developments [11,12], many new ADDIs are discovered by accident as the events of ADDI happen unexpectedly, which delays the treatments, aggravates the disease conditions, and even results in significant morbidity and mortality [13–15]. The occurrences of undetected ADDIs mainly lie on the elaborate chemical reactions among different kinds of drugs and the complex interactions between drugs and human body [16], which make it hard for the vitro experiments to sufficiently simulate the vivo environments so as to detect potential ADDIs [17]. Thus, it brings a great urgency to predict ADDIs for preventing this life-threatening issue and providing safe treatments in clinical practice [10,18].

In the past few years, laboratory-based methods (e.g., testing cytochrome P450 or transporter-associated pharmacokinetic interactions [11], analyzing drug metabolites [19], and modeling the bindings between drugs and receptors [20], etc.) have been developed for ADDI prediction, which are restricted in terms of poor statistical significance, tremendous research expenditure as well as high time consumption [5,9]. As the availability of healthcare data increases, the feasibility of data-driven models has been broadly explored by facilitating knowledge-based methods for ADDI prediction [8]. To be specific, researchers proposed to adopt text mining [21], data mining [22], and semantic analysis techniques [23] to predict ADDIs from post-event materials, such as biomedical text [21], electronic medical records [24], and the FDA adverse event reporting system (FAERS) [25]. The success of these

* Corresponding author.
E-mail address: liuyg@uestc.edu.cn (Y. Liu).

https://doi.org/10.1016/j.knosys.2020.105978
0950-7051/

methods primarily lies on the importance of pharmacovigilance in an increasing attention and the accumulation of healthcare data in a massive scale [26]. However, due to that knowledge-based methods are highly dependent on the sufficient clinical evidences, their predictive results can only be regarded as a post-marketing surveillance, serving as post-marketing warnings, but are inadequate in exploring potential ADDIs before drugs hit markets, serving as pre-marketing alerts [10,18].

Based on the assumption that ADDIs are essentially caused by the attribute conflicts between two drugs [19,27], similarity-based methods are proposed to provide such alerts for predicting potential ADDIs. These methods first extract drug attributes (e.g., molecular structure, phenotype, target, pharmacokinetics, and pathway, etc.) from medical databases (e.g., DrugBank [28], PubChem [29], and KEGG [30], etc.), and then compute similarity scores among drugs used as inputs for machine learning classifiers [31–33]. In particular, molecular structure is the most widely used drug attribute in the predictive models, since the incompatibility of molecular structure is a critical factor that leads to many ADDIs [5,14,34]. Meanwhile, recent researches also show that integrating multiple drug attributes into the same model can enhance the prediction performance [33,35]. On the whole, the advantage of similarity-based methods lies on that they mainly focus on the biochemical data of drugs from medical databases rather than post-event materials, which benefits drug safety experts to predict potential ADDIs in advance [31,36]. As a result, their prediction results can reduce investigative scope for laboratory-based methods, provide appropriate drug regulatory for healthcare system, and ultimately improve medication safety for clinical treatments [8,37].

Although the impressive results are achieved by the above-mentioned approaches, there are still some limitations:

- Most methods only focus on whether an adverse interaction would happen or not for a drug pair. In practice, there are multiple specific adverse interactions that are associated with a particular adverse drug pair [25,36]. For example, the co-administration of *Aspirin* and *Warfarin* can cause gastrointestinal hemorrhage, excessive bleeding, and urethrorrhagia [3], and the combination of *Zidovudine* and *Acyclovir* can result in nausea, vomiting, renalinsufficiency, and lethargy [38]. Furthermore, the prediction of specific types of adverse interactions can uncover the underlying threats behind each adverse drug pair, which is helpful to provide positive measures for the prevention of ADDIs [18, 35].
- They typically adopt biochemical-based attributes, such as molecular structure similarity, interaction profile fingerprints, and pharmacokinetic interactions for ADDI prediction [14,33,36]. However, many ADDIs are essentially caused by the aggravation of side effects. More specifically, the interactions between two drugs may worsen their respective side effects and result in severe ADDIs [7,15,37]. For example, the co-administration of *Aspirin* and *Warfarin* may aggravate the gastrointestinal ulcer caused by *Aspirin* and worsen the intestinal wall edema induced by *Warfarin*, which finally increases the risks of gastrointestinal hemorrhage induced by their interaction [3]. The combination of *Zidovudine* and *Acyclovir* will aggravate the flatulence induced by *Zidovudine* and worsen the side effect of abdominal discomfort caused by *Acyclovir*, which ultimately enhances the dangers of nausea and vomiting [39].
- Most importantly, the current models usually work in a "black-box" manner for ADDI prediction, i.e., they only focus on computing attribute similarities to measure the incompatibility among drugs [14,40], or constructing similarity networks to infer the unknown ADDIs [9,34], or designing neural networks to capture the underlying mappings between attributes and adverse interactions [6,31]. However, these methods are insufficient in exploring the potential adverse mechanisms among drugs and incapable of revealing the leading factors of ADDIs. To our best knowledge, no previous work has incorporated the exploration of adverse mechanisms and leading factors into ADDI prediction.

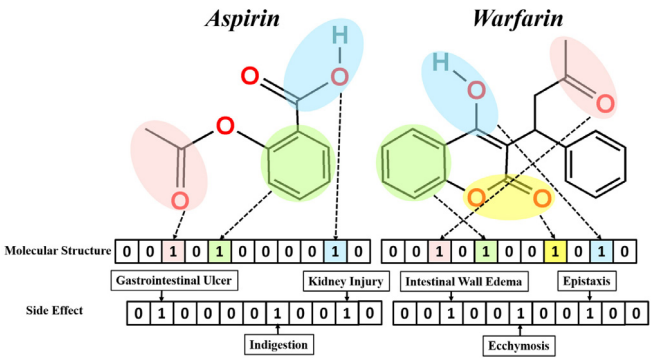


Fig. 1. The schematic of molecular structure and side effect information for *Aspirin* and *Warfarin*, respectively. The values of the entries (1s or 0s) indicate whether a drug owns the corresponding molecular substructures and side effects or not.

To address above issues, we propose a novel Multi-Task Multi-Attribute (MTMA) learning model for ADDI prediction. In MTMA, we regard the task of ADDI prediction as a multi-task learning problem and adopt two drug attributes, molecular structure and side effect, to model the adverse interactions among drugs. Fig. 1 presents the schematic of molecular structure and side effect information for *Aspirin* and *Warfarin*, respectively. To investigate the adverse mechanisms among drugs, we design two interpretable tensors, adverse molecular structure–molecular structure interaction tensor and adverse side effect-side effect interaction tensor, to capture the relationships between attributes and adverse interactions. Due to that the biological activities of a drug are usually related to a few privileged molecular substructures [41,42] and the adverse reactions caused by a drug are often associated with a few significant side effects [43,44], we impose $l_{2,1}$-norm on the predicted attribute matrices to explore the leading molecular substructures and side effects for each specific ADDI. This regularization is also critical for revealing the potential adverse mechanisms among drugs and can improve its prediction results. To solve the optimization problem of MTMA, we develop an efficient alternating algorithm based on the methods of low-rank tensor decomposition [45] and stochastic gradient descent (SGD) [46]. Experimental results on the real-world dataset demonstrate the effectiveness of MTMA when compared with nine baseline methods and its three variants. By conducting a case study for visualized presentation, we find that MTMA is capable of uncovering the adverse mechanisms and leading factors of ADDIs for benefiting drug safety studies.

The rest of paper is organized as follows. Section 2 reviews some related work on ADDI prediction. In Section 3, we introduce the proposed MTMA model. Section 4 provides the optimization algorithm of MTMA. Experimental results are reported in Section 5. Finally, Section 6 provides conclusions and future work.

2. Related work

ADDI prediction is an important topic in the field of ADR research. For knowledge-based methods, Cai et al. [4] introduced a causal concept into association rule mining and proposed a causal association rule by exploiting the properties of

V-structures in Bayesian networks to predict ADDIs from FAERS. Tatonetti et al. [8] proposed a statistical correction of uncharacterized bias approach to remove synthetic associations, coprescriptions, and hidden covariates from prescriptions for promoting ADDI prediction. Jiang et al. [23] developed a semantic web-based approach and utilized a normalized FAERS dataset to detect severe ADDIs from adverse drug events. By partitioning ADDIs into five groups based on their syntactic structures, Bui et al. [21] proposed a feature-based approach to map ADDIs into a suitable syntactic structure so as to detect ADDIs from biomedical text. Shen et al. [26] developed a knowledge oriented feature-driven method to learn drug-related knowledge for ADDI prediction and similarity computation.

Recently, with the increasing attention of deep learning techniques, researchers proposed to adopt convolutional neural network (CNN) [47], recurrent neural network (RNN) [48], and long short term memory neural network (LSTM) [49] to detect ADDIs from biomedical literature, social media, and heterogeneous databases. Based on the method of syntax word embedding, Zhao et al. [47] proposed a syntax CNN that employs the syntactic information of sentences for extracting ADDIs from biomedical literature. To label ADDIs in tweet posts, Cocos et al. [48] developed a RNN with bidirectional LSTM to process tweet posts in forward and backward directions for ADDI prediction. In [49], Huang et al. first used a feature-based binary classifier to identify the positive instances and then developed an LSTM-based model to classify the positive instances into specific category, which is effective in the detection of ADDIs from biomedical literature.

To address the problem of knowledge-based methods in predicting potential ADDIs in advance, similarity-based methods are proposed with the extensive studies on biochemical characteristics of drugs [5,6,14,31,37,40]. Vilar et al. [5,14] first adopted interaction profile fingerprints to measure the similarities among drugs, and then proposed a large-scale model to predict unknown ADDIs from known ADDIs. By learning the representation of interactions between drug pairs and characterizing their complicated non-linear interactions, Chu et al. [6] developed a multi-label robust factorization autoencoder for ADDI prediction. By incorporating molecular structure and drug name as inputs, Ryu et al. [31] designed a deep learning framework for predicting ADDIs associated with 191,878 drug pairs. Shi et al. [37] designed a triple matrix factorization-based unified framework by leveraging the side effects of drugs to predict both binary ADDIs and comprehensive ADDIs under different screening scenarios, which is effective in exploring the pharmacological changes caused by the interactions among drugs. Based on the functional information of drugs, Ferdousi et al. [40] used inner product to measure the functional similarities among drugs for predicting ADDIs associated with 2189 approved drugs.

In the past years, many research efforts are also extensively made by incorporating multiple drug attributes into ADDI prediction, which achieve considerable performance [27,32,33,36]. For instance, Kastrin et al. [27] designed a statistical learning method to predict ADDIs based on the topological and semantic similarities among drugs. Zhang et al. [32] first developed a sparse feature learning method to map different drug attribute spaces into a common interaction space, and then introduced a linear neighborhood regularization to model the adverse interactions among drugs in the interaction space for ADDI prediction. Based on phenotypic, therapeutic, molecular structure, and genomic similarities among drugs, Cheng and Zhao [33] applied five machine learning algorithms, i.e., naive Bayes, decision tree, k-nearest neighbors, logistic regression, and support vector machine, in the heterogeneous network-assisted inference framework to identify potential ADDIs during clinical trials. Zhang et al. [36] proposed a manifold regularized matrix factorization model to predict potential ADDIs by introducing drug attribute-based manifold regularization into matrix factorization.

However, the abovementioned researches only focus on binary predictions that cannot identify the specific adverse interactions associated with a particular drug pair. Meanwhile, multi-task learning is a popular learning paradigm in machine learning that aims to exploit useful information included in multiple related tasks for improving their generalization in model applications [50–53]. To this end, many multi-task learning models are recently introduced by regarding each specific ADDI as a learning task for early ADDI prediction [18,35,54]. For example, Jin et al. [18] designed a multi-task dyadic regression method to detect specific types of ADDIs, in which the detection for each ADDI is regarded as a regression task. Shi et al. [35] proposed a balance regularized semi-nonnegative matrix factorization method to partition drugs into communities and predict comprehensive ADDIs in a cold-start scenario.

Different from the current methods, in this paper, we aim to reveal the potential adverse mechanisms and explore the leading factors of ADDIs by employing two drug attributes, designing two interpretable tensors, and incorporating $l_{2,1}$-norm regularization for ADDI prediction.

3. The proposed MTMA model

3.1. The basic prediction model

Let $D = \{d_1, d_2, \ldots, d_N\}$ be a set of N drugs. Some pairs of drugs are detected as adverse drug pairs that result in multiple adverse interactions. Suppose that there are K different adverse interactions. We use tensor $\mathcal{R} \in \{0, 1\}^{N\times N\times K}$ to represent the adverse interactions among drugs such that if drug pair (d_i, d_j) is associated with the kth adverse interaction, then $\mathcal{R}_{i,j,k} = 1$; otherwise $\mathcal{R}_{i,j,k} = 0$. In MTMA, the prediction of each specific ADDI is regarded a classification task and the adverse interactions among drugs are assumed to follow a Gaussian distribution, which can be formulated as follows:

$$p(\mathcal{R} \mid \widetilde{\mathcal{R}}, \sigma^2) = \prod_{i,j,k} [\mathcal{N}(\mathcal{R}_{i,j,k} | \widetilde{\mathcal{R}}_{i,j,k}, \sigma^2)]^{\mathcal{R}_{i,j,k}}, \tag{1}$$

where $\widetilde{\mathcal{R}} \in \mathbb{R}^{N\times N\times K}$ is the estimated tensor of $\mathcal{R}$ with its entry $\widetilde{\mathcal{R}}_{i,j,k}$ being the estimated value of $\mathcal{R}_{i,j,k}$, $\mathcal{N}(x|\mu, \sigma^2)$ is the probability density function of the Gaussian distribution with mean μ and variance σ^2. Due to that ADDIs are essentially caused by the incompatibility of molecular structures and the aggravation of side effects [7,14], we adopt two drug attributes, molecular structure and side effect, to model the adverse interactions among drugs. Each drug is associated with an M-dimensional molecular structure space and a J-dimensional side effect space. For drug d_i, we use $\mathbf{P}_i \in \mathbb{R}^{1\times M}$ and $\mathbf{Q}_i \in \mathbb{R}^{1\times J}$ to denote its predicted molecular structure and side effect vectors, respectively. In this way, the estimated ADDI value $\widetilde{\mathcal{R}}_{i,j,k}$ can be calculated as follows:

$$\widetilde{\mathcal{R}}_{i,j,k} = \widetilde{\mathcal{R}}^{(1)}_{i,j,k} + \widetilde{\mathcal{R}}^{(2)}_{i,j,k}, \tag{2}$$

where $\widetilde{\mathcal{R}}^{(1)}_{i,j,k}$ and $\widetilde{\mathcal{R}}^{(2)}_{i,j,k}$ represent the estimated adverse values contributed from molecular structure and side effect, respectively.[1]

[1] In this paper, $\widetilde{\mathcal{R}}^{(1)}_{i,j,k}$ and $\widetilde{\mathcal{R}}^{(2)}_{i,j,k}$ are integrated through an addition operation to formulate $\widetilde{\mathcal{R}}_{i,j,k}$, in which the adverse values contributed from molecular structure and side effect are under the same weight. Actually, we can assign different weights to $\widetilde{\mathcal{R}}^{(1)}_{i,j,k}$ and $\widetilde{\mathcal{R}}^{(2)}_{i,j,k}$ for formulating $\widetilde{\mathcal{R}}_{i,j,k}$, e.g., $\widetilde{\mathcal{R}}_{i,j,k} = r_1\widetilde{\mathcal{R}}^{(1)}_{i,j,k} + r_2\widetilde{\mathcal{R}}^{(2)}_{i,j,k}$. In this way, the importance of molecular structure and side effect on ADDI modeling can be learned by MTMA. Additionally, more complicated integration strategies such as stacking, boosting and neural network methods, can also be applied for learning the weights of different attributes, which is the focus of our future work.

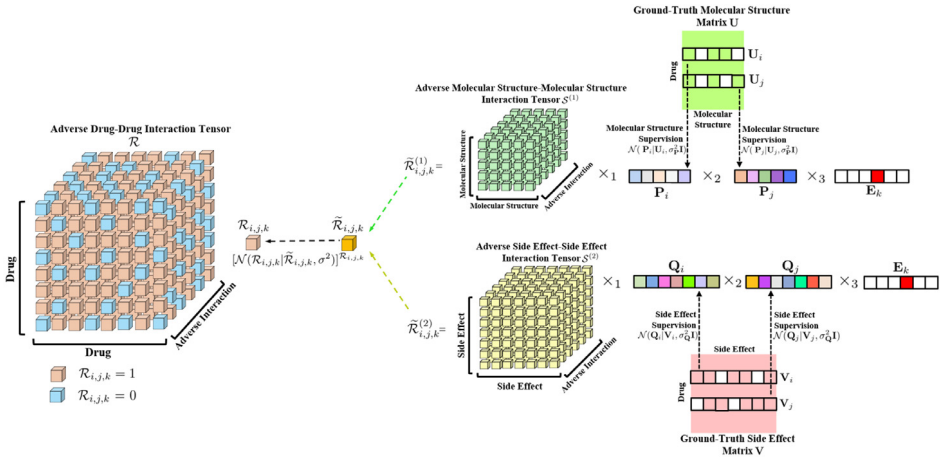


Fig. 2. The framework of the proposed MTMA model for ADDI prediction.

In particular,

$$\begin{aligned}\widetilde{\mathcal{R}}^{(1)}_{i,j,k} &= \mathcal{S}^{(1)} \times_1 \mathbf{P}_i \times_2 \mathbf{P}_j \times_3 \mathbf{E}_k,\\ \widetilde{\mathcal{R}}^{(2)}_{i,j,k} &= \mathcal{S}^{(2)} \times_1 \mathbf{Q}_i \times_2 \mathbf{Q}_j \times_3 \mathbf{E}_k,\end{aligned} \tag{3}$$

where $\mathcal{S}^{(1)} \in \mathbb{R}^{M\times M\times K}$ denotes the adverse molecular structure–molecular structure interaction tensor with its entry $\mathcal{S}^{(1)}_{i,j,k}$ being the adverse relation between the ith molecular substructure and the jth molecular substructure with respect to the kth adverse interaction, $\mathcal{S}^{(2)} \in \mathbb{R}^{J\times J\times K}$ represents the adverse side effect-side effect interaction tensor with its entry $\mathcal{S}^{(2)}_{i,j,k}$ being the adverse relation between the ith side effect and the jth side effect with respect to the kth adverse interaction. $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ are two interpretable tensors that are used to capture the relationships between attributes and adverse interactions so as to reveal the adverse mechanisms among drugs. $\times_k$ denotes the mode-k product of a tensor and $\mathbf{E}_k$ represents the one-hot vector with the length of K and the kth entry being 1.

To solve the difficulties in estimating drug attributes and reasonably reveal the adverse mechanisms as well as enhance the prediction performance, we place spherical Gaussian priors on the predicted molecular structure vector $\mathbf{P}_i$ and side effect vector $\mathbf{Q}_i$ under the supervision of the ground-truth molecular structure vector $\mathbf{U}_i \in \mathbb{R}^{1\times M}$ and side effect vector $\mathbf{V}_i \in \mathbb{R}^{1\times J}$, respectively, which can be formulated as follows:

$$\begin{aligned}p(\mathbf{P}|\mathbf{U},\sigma^2_{\mathbf{P}}) &= \prod_{i=1}^{N} \mathcal{N}(\mathbf{P}_i|\mathbf{U}_i,\sigma^2_{\mathbf{P}}\mathbf{I}),\\ p(\mathbf{Q}|\mathbf{V},\sigma^2_{\mathbf{Q}}) &= \prod_{i=1}^{N} \mathcal{N}(\mathbf{Q}_i|\mathbf{V}_i,\sigma^2_{\mathbf{Q}}\mathbf{I}),\end{aligned} \tag{4}$$

where $\mathbf{P} \in \mathbb{R}^{N\times M}$ and $\mathbf{Q} \in \mathbb{R}^{N\times J}$ are the predicted molecular structure and side effect matrices of drugs, respectively, $\mathbf{U} \in \mathbb{R}^{N\times M}$ and $\mathbf{V} \in \mathbb{R}^{N\times J}$ are the ground-truth molecular structure and side effect matrices of drugs, respectively, $\sigma^2_{\mathbf{P}}$ and $\sigma^2_{\mathbf{Q}}$ are denoted as the variance parameters, $\mathbf{I}$ is the identity matrix. Based on the Bayesian inference, we derive the following equation:

$$p(\mathbf{P},\mathbf{Q},\mathcal{S}^{(1)},\mathcal{S}^{(2)} \mid \mathcal{R},\ \mathbf{U},\ \mathbf{V},\sigma^2,\sigma^2_{\mathbf{P}},\sigma^2_{\mathbf{Q}}) \propto p(\mathcal{R}\mid\widetilde{\mathcal{R}},\sigma^2)p(\mathbf{P}|\mathbf{U},\sigma^2_{\mathbf{P}})p(\mathbf{Q}|\mathbf{V},\sigma^2_{\mathbf{Q}}). \tag{5}$$

Fig. 2 presents the framework of the proposed MTMA model for ADDI prediction.[2] The model parameters (i.e., $\mathbf{P}$, $\mathbf{Q}$, $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$) in MTMA can be learned by maximizing the log-posterior of the objective function in Eq. (5), which is equivalent to solving the following sum-of-squared-errors objective function with quadratic regularization terms:

$$\min_{\mathbf{P},\mathbf{Q},\mathcal{S}^{(1)},\mathcal{S}^{(2)}} \frac{1}{2\sigma^2}\|\mathcal{R}\odot(\widetilde{\mathcal{R}}-\mathcal{R})\|^2_F + \frac{1}{2\sigma^2_{\mathbf{P}}}\|\mathbf{P}-\mathbf{U}\|^2_F + \frac{1}{2\sigma^2_{\mathbf{Q}}}\|\mathbf{Q}-\mathbf{V}\|^2_F, \tag{6}$$

where $\odot$ denotes the Hadamard product of two tensors, and $\|\cdot\|$ represents the Frobenius norm of a tensor or a matrix. In this paper, we set $\sigma^2 = \sigma^2_{\mathbf{P}} = \sigma^2_{\mathbf{Q}}$ for the convenience of formulation derivations.

3.2. Self-representation of drug attributes

In the former subsection, we develop a multi-task learning model to formulate the problem of ADDI prediction under the supervision of the ground-truth attribute information (see Eq. (6)). In practice, drugs are usually connected with each other through different attribute relationships and the attributes of a drug do not deviate significantly from the majority of other drugs [55]. One natural strategy can be adopted to reconstruct the predicted attribute matrices by some representative drugs. To be specific, the predicted attribute vectors of a drug can be represented by the vectors of some representative drugs, which means that the predicted attribute matrices can be self-represented by the attribute information of these representative drugs. Mathematically, the self-representation of drug attributes can be formulated as follows:

$$\min_{\mathbf{P},\mathbf{Q},\mathbf{W}} \|\mathbf{P}-\mathbf{W}^{\mathrm{T}}\mathbf{P}\|^2_F + \|\mathbf{Q}-\mathbf{W}^{\mathrm{T}}\mathbf{Q}\|^2_F + \beta\|\mathbf{W}\|_{2,0}, \tag{7}$$

where $\mathbf{W} \in \mathbb{R}^{N\times N}$ is denoted as the coefficient matrix such that the predicted molecular structure and side effect vector for each drug can be represented by a linear combination of other drugs, $\cdot^{\mathrm{T}}$ and $\|\cdot\|_{2,0}$ represent the transpose and $l_{2,0}$-norm of a matrix, respectively, $\|\mathbf{W}\|_{2,0}$ is used to ensure that only the attribute vectors of a few representative drugs can be employed to recover the predicted attribute matrices, β is the regularization parameter

[2] In practice, MTMA can be extended to a more generalized one by simultaneously taking other drug attributes into consideration. Concisely, we can generalize Eq. (2) into $\widetilde{\mathcal{R}}_{i,j,k} = \sum_{t=1}^{T}\widetilde{\mathcal{R}}^{(t)}_{i,j,k}$, where T is the number of attributes to be considered and $\widetilde{\mathcal{R}}^{(t)}_{i,j,k}$ represents the estimated adverse value contributed from the tth attribute. Similar to Eq. (3), $\widetilde{\mathcal{R}}^{(t)}_{i,j,k}$ might also be computed based on the corresponding tensor and the attribute vectors of drugs. Due to that molecular structure and side effect are essential factors in ADDI modeling, these two attributes are adopted in this paper for ADDI prediction. In the future work, we will incorporate more attributes into ADDI prediction and analyze the effects of different attributes on ADDI modeling.

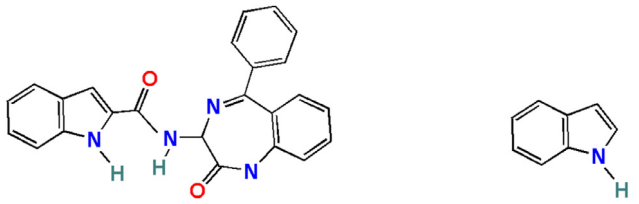


Fig. 3. The molecular structure of *Devazepide* and its privileged molecular substructure [41].

that controls the row sparsity of matrix **W**. However, the problem in Eq. (7) is NP-hard due to the existence of $l_{2,0}$-norm. Since $\|\mathbf{W}\|_{2,1}$ is the minimum convex hull of $\|\mathbf{W}\|_{2,0}$, the minimization of $\|\mathbf{W}\|_{2,0}$ is equivalent to the minimization of $\|\mathbf{W}\|_{2,1}$, which is widely used in feature selection task [56,57]. In this way, Eq. (7) can be rewritten in the following:

$$\min_{\mathbf{P},\mathbf{Q},\mathbf{W}} \|\mathbf{P}-\mathbf{W}^{\mathrm{T}}\mathbf{P}\|_F^2+\|\mathbf{Q}-\mathbf{W}^{\mathrm{T}}\mathbf{Q}\|_F^2+\beta\|\mathbf{W}\|_{2,1}. \tag{8}$$

3.3. Exploration of leading factors

Recent studies show that the biological activities of a drug are usually related to a few privileged molecular substructures [41,42]. Fig. 3 shows the molecular structure of *Devazepide* (a cholecystokinin antagonist used to treat gastrointestinal disorders) and its privileged molecular substructure. To be specific, the privileged molecular substructure in the right of Fig. 3 is highly related to the biological activity of *Devazepide* in binding to opioid receptors and other protein targets [41]. Fig. 4 presents three privileged molecular substructures of drugs. According to the results shown in [42], the drugs with these privileged molecular substructures usually own the biological activity of anti-sickling in disease treatments.

Meanwhile, in terms of the side effects of drugs, recent researches indicate that the adverse reactions caused by a drug are often associated with a few significant side effects [43,44]. Table 1 provides the side effects and significant side effects of *Aspirin*, *Warfarin*, *Zidovudine*, and *Acyclovir*, respectively [58]. Taking *Aspirin* as an example, we find that *Aspirin* can result in multiple side effects, but its adverse reactions are primarily associated with its three significant side effects (i.e., gastrointestinal ulcer, indigestion, and kidney injury), while the remaining side effects rarely happen or usually accompany with the significant ones in most cases.

Based on the abovementioned results, we can infer that the adverse interactions among drugs are essentially caused by the incompatibility of a few privileged molecular substructures and the aggravation of a few significant side effects, which inspires us to explore the leading molecular substructures and side effects for each specific ADDI so as to reveal the adverse mechanisms among drugs and benefit ADDI prediction results. As $\mathbf{P}\in\mathbb{R}^{N\times M}$ and $\mathbf{Q}\in\mathbb{R}^{N\times J}$ are denoted as the predicted molecular structure and side effect matrices of drugs, respectively, we impose $l_{2,1}$-norm on matrices **P** and **Q** for constraining their row sparsity so as to explore the leading molecular substructures and side effects for each ADDI, which can be mathematically formulated as follows:

$$\min_{\mathbf{P},\mathbf{Q}} \|\mathbf{P}\|_{2,1}+\|\mathbf{Q}\|_{2,1}. \tag{9}$$

3.4. Objective function of MTMA

By incorporating Eqs. (8) and (9) into Eq. (6), we obtain the final objective function of the proposed MTMA model as follows:

$$\begin{aligned}\min_{\mathbf{P},\mathbf{Q},\mathbf{W},\mathcal{S}^{(1)},\mathcal{S}^{(2)}} &\frac{1}{2\sigma^2}\|\mathcal{R}\odot(\widetilde{\mathcal{R}}-\mathcal{R})\|_F^2+\frac{1}{2\sigma^2}\|\mathbf{P}-\mathbf{U}\|_F^2+\frac{1}{2\sigma^2}\|\mathbf{Q}-\mathbf{V}\|_F^2\\ &+\frac{\alpha}{2}(\|\mathbf{P}-\mathbf{W}^{\mathrm{T}}\mathbf{P}\|_F^2+\|\mathbf{Q}-\mathbf{W}^{\mathrm{T}}\mathbf{Q}\|_F^2)+\beta(\|\mathbf{P}\|_{2,1}+\|\mathbf{Q}\|_{2,1}+\|\mathbf{W}\|_{2,1}),\end{aligned} \tag{10}$$

where parameter α controls the self-representation of the predicted molecular structure and side effect matrices, parameter β controls the row sparsity of matrices **P**, **Q**, and **W**.

4. Optimization algorithm

The objective function of MTMA in Eq. (10) cannot be optimized directly due to the existence of tensors $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$. Meanwhile, it is also not convex with respect to matrices **P**, **Q**, and **W**, simultaneously. In this section, we first adopt low-rank tensor decomposition [45] to decompose tensors $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$, then develop an efficient alternating algorithm based on SGD [46] to solve the optimization problem of MTMA.

4.1. Low-rank tensor decomposition

In this paper, we assume that the high-order tensors $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ are low rank, which can be approximated by the product of a set of low rank matrices. Here, low-rank CANDECOMP/PARAFAC (CP) decomposition method [45] is adopted to factorize tensors $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ into a sum of component rank-one tensors:

$$\begin{aligned}\mathcal{S}^{(1)}&\approx\sum_{l=1}^{L}\mathbf{A}_l\circ\mathbf{A}_l\circ\mathbf{C}_l,\\ \mathcal{S}^{(2)}&\approx\sum_{l=1}^{L}\mathbf{B}_l\circ\mathbf{B}_l\circ\mathbf{C}_l,\end{aligned} \tag{11}$$

where L is the rank of the tensors, $L\ll M,J,K$, $\circ$ is the outer product of vectors, $\mathbf{A}\in\mathbb{R}^{L\times M}$, $\mathbf{B}\in\mathbb{R}^{L\times J}$, and $\mathbf{C}\in\mathbb{R}^{L\times K}$ denote the latent representations of molecular structure, side effect, and adverse interaction, respectively. By substituting Eq. (11) back to Eq. (3), we rewrite $\widetilde{\mathcal{R}}^{(1)}_{i,j,k}$ and $\widetilde{\mathcal{R}}^{(2)}_{i,j,k}$ as follows:

$$\begin{aligned}\widetilde{\mathcal{R}}^{(1)}_{i,j,k}&\approx\sum_{l=1}^{L}(\mathbf{A}_l\mathbf{P}_i^{\mathrm{T}})(\mathbf{A}_l\mathbf{P}_j^{\mathrm{T}})(\mathbf{C}_l\mathbf{E}_k^{\mathrm{T}}),\\ \widetilde{\mathcal{R}}^{(2)}_{i,j,k}&\approx\sum_{l=1}^{L}(\mathbf{B}_l\mathbf{Q}_i^{\mathrm{T}})(\mathbf{B}_l\mathbf{Q}_j^{\mathrm{T}})(\mathbf{C}_l\mathbf{E}_k^{\mathrm{T}}).\end{aligned} \tag{12}$$

4.2. Parameter update

In this subsection, we first fix matrix **W** to update matrices **P**, **Q**, **A**, **B**, and **C**, and then fix **P**, **Q**, **A**, **B**, and **C** to update **W**.

4.2.1. Update **P**, **Q**, **A**, **B**, and **C**

When matrix **W** is fixed, the objective function with respect to matrices **P**, **Q**, **A**, **B**, and **C** can be written as

$$\begin{aligned}\min_{\mathbf{P},\mathbf{Q},\mathbf{A},\mathbf{B},\mathbf{C}} &\frac{1}{2\sigma^2}\|\mathcal{R}\odot(\widetilde{\mathcal{R}}-\mathcal{R})\|_F^2+\frac{1}{2\sigma^2}\|\mathbf{P}-\mathbf{U}\|_F^2+\frac{1}{2\sigma^2}\|\mathbf{Q}-\mathbf{V}\|_F^2\\ &+\frac{\alpha}{2}(\|\mathbf{P}-\mathbf{W}^{\mathrm{T}}\mathbf{P}\|_F^2+\|\mathbf{Q}-\mathbf{W}^{\mathrm{T}}\mathbf{Q}\|_F^2)+\beta(\|\mathbf{P}\|_{2,1}+\|\mathbf{Q}\|_{2,1}).\end{aligned} \tag{13}$$

To solve the optimization problem in Eq. (13), we adopt SGD [46] to derive the update rules of model parameters. At each iteration, we randomly select one sample $(d_i, d_j, k, \mathcal{R}_{i,j,k})$ from the training

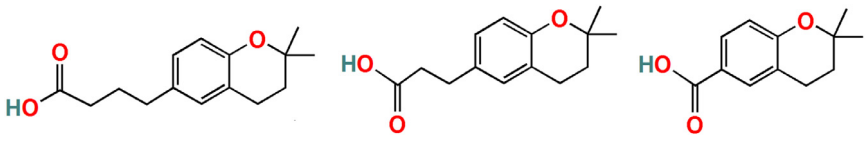


Fig. 4. Three privileged molecular substructures of drugs [42].

Table 1
The side effects and the significant side effects of *Aspirin*, *Warfarin*, *Zidovudine*, and *Acyclovir*, respectively [58]. The significant side effects for each drug are in bold.

Drug	Side effect
Aspirin	**gastrointestinal ulcer**, **indigestion**, **kidney injury**, respiratory alkalosis, hepatotoxicity, pruritus, diuresis, convulsion, leukopenia, thrombocytopenia, and etc.
Warfarin	**intestinal wall edema**, **ecchymosis**, **epistaxis**, purpura, breast necrosis, microangiopathy, cutaneous necrosis, intestinal obstruction, cutaneous gangrene, and etc.
Zidovudine	**flatulence**, **hypoxemia**, **pectoralgia**, dysphagia, myalgia, acidosis, abdominal cramps, bacteraemia, cystitis, cardiac disorder, haematuria, diarrhoea, and etc.
Acyclovir	**abdominal discomfort**, **somnolence**, **hematuria**, anaemia, leukopenia, vasculitis, oedema, malaise diarrhoea, erythema, renal failure, tachycardia, stupor, hypokalaemia, and etc.

set to update $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. Based on the formulations in Eqs. (2) and (12), we first compute the partial derivatives of $\widetilde{\mathcal{R}}_{i,j,k}$ with respect to vectors $\mathbf{P}_i$, $\mathbf{Q}_j$ and matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ as follows:

$$\begin{aligned}
\frac{\partial \widetilde{\mathcal{R}}_{i,j,k}}{\partial \mathbf{P}_i} &= \frac{\partial \widetilde{\mathcal{R}}^{(1)}_{i,j,k}}{\partial \mathbf{P}_i} = \sum_{l=1}^{L} \mathbf{A}_l (\mathbf{A}_l \mathbf{P}_j^{\mathrm{T}})(\mathbf{C}_l \mathbf{E}_k^{\mathrm{T}}),\\
\frac{\partial \widetilde{\mathcal{R}}_{i,j,k}}{\partial \mathbf{Q}_j} &= \frac{\partial \widetilde{\mathcal{R}}^{(2)}_{i,j,k}}{\partial \mathbf{Q}_j} = \sum_{l=1}^{L} \mathbf{B}_l (\mathbf{B}_l \mathbf{Q}_i^{\mathrm{T}})(\mathbf{C}_l \mathbf{E}_k^{\mathrm{T}}),\\
\frac{\partial \widetilde{\mathcal{R}}_{i,j,k}}{\partial \mathbf{A}} &= \frac{\partial \widetilde{\mathcal{R}}^{(1)}_{i,j,k}}{\partial \mathbf{A}} = diag[(\mathbf{A}\mathbf{P}_j^{\mathrm{T}})(\mathbf{E}_k \mathbf{C}^{\mathrm{T}})]\mathbf{P}_i + diag[(\mathbf{A}\mathbf{P}_i^{\mathrm{T}})(\mathbf{E}_k \mathbf{C}^{\mathrm{T}})]\mathbf{P}_j,\\
\frac{\partial \widetilde{\mathcal{R}}_{i,j,k}}{\partial \mathbf{B}} &= \frac{\partial \widetilde{\mathcal{R}}^{(2)}_{i,j,k}}{\partial \mathbf{B}} = diag[(\mathbf{B}\mathbf{Q}_j^{\mathrm{T}})(\mathbf{E}_k \mathbf{C}^{\mathrm{T}})]\mathbf{Q}_i + diag[(\mathbf{B}\mathbf{Q}_i^{\mathrm{T}})(\mathbf{E}_k \mathbf{C}^{\mathrm{T}})]\mathbf{Q}_j,\\
\frac{\partial \widetilde{\mathcal{R}}_{i,j,k}}{\partial \mathbf{C}} &= diag[(\mathbf{A}\mathbf{P}_i^{\mathrm{T}})(\mathbf{P}_j \mathbf{A}^{\mathrm{T}})]\mathbf{E}_k + diag[(\mathbf{B}\mathbf{Q}_i^{\mathrm{T}})(\mathbf{Q}_j \mathbf{B}^{\mathrm{T}})]\mathbf{E}_k,
\end{aligned} \tag{14}$$

where $diag(\cdot)$ denotes a column vector with its entries being the diagonal elements of a matrix. Then, we compute the partial derivatives of Eq. (13) with respect to $\mathbf{P}_i$, $\mathbf{Q}_j$, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ and write their update rules as follows:

$$\begin{aligned}
\mathbf{P}_i &\leftarrow \mathbf{P}_i + \lambda[\frac{\Delta_{i,j,k}}{\sigma^2}\frac{\partial \widetilde{\mathcal{R}}^{(1)}_{i,j,k}}{\partial \mathbf{P}_i} - \frac{1}{\sigma^2}(\mathbf{P}_i - \mathbf{U}_i) - \alpha \boldsymbol{\Theta}_i \mathbf{P} - \beta(\mathbf{P}_i \mathbf{P}_i^{\mathrm{T}})^{-\frac{1}{2}}\mathbf{P}_i],\\
\mathbf{Q}_j &\leftarrow \mathbf{Q}_j + \lambda[\frac{\Delta_{i,j,k}}{\sigma^2}\frac{\partial \widetilde{\mathcal{R}}^{(2)}_{i,j,k}}{\partial \mathbf{Q}_j} - \frac{1}{\sigma^2}(\mathbf{Q}_j - \mathbf{V}_j) - \alpha \boldsymbol{\Theta}_j \mathbf{Q} - \beta(\mathbf{Q}_j \mathbf{Q}_j^{\mathrm{T}})^{-\frac{1}{2}}\mathbf{Q}_j],\\
\mathbf{A} &\leftarrow \mathbf{A} + \lambda\frac{\Delta_{i,j,k}}{\sigma^2}\frac{\partial \widetilde{\mathcal{R}}^{(1)}_{i,j,k}}{\partial \mathbf{A}}, \mathbf{B} \leftarrow \mathbf{B} + \lambda\frac{\Delta_{i,j,k}}{\sigma^2}\frac{\partial \widetilde{\mathcal{R}}^{(2)}_{i,j,k}}{\partial \mathbf{B}},\\
\mathbf{C} &\leftarrow \mathbf{C} + \lambda\frac{\Delta_{i,j,k}}{\sigma^2}\frac{\partial \widetilde{\mathcal{R}}_{i,j,k}}{\partial \mathbf{C}},
\end{aligned} \tag{15}$$

where $\Delta_{i,j,k} = \mathcal{R}_{i,j,k} - \widetilde{\mathcal{R}}_{i,j,k}$, $\boldsymbol{\Theta} = (\mathbf{I} - \mathbf{W})(\mathbf{I} - \mathbf{W})^{\mathrm{T}}$, and λ is the learning rate.

4.2.2. Update $\mathbf{W}$

To derive the update rule for matrix $\mathbf{W}$, we fix matrices $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ and remove the terms that are irrelevant to $\mathbf{W}$. Thus, the objective function with respect to matrix $\mathbf{W}$ can be written as

$$\min_{\mathbf{W}} \frac{\alpha}{2}(\|\mathbf{P} - \mathbf{W}^{\mathrm{T}}\mathbf{P}\|_F^2 + \|\mathbf{Q} - \mathbf{W}^{\mathrm{T}}\mathbf{Q}\|_F^2) + \beta\|\mathbf{W}\|_{2,1}. \tag{16}$$

By calculating the partial derivatives of Eq. (16) with respect to $\mathbf{W}$ and setting it to 0, we have

$$\alpha(\mathbf{P}\mathbf{P}^{\mathrm{T}}\mathbf{W} + \mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{W} - \mathbf{P}\mathbf{P}^{\mathrm{T}} - \mathbf{Q}\mathbf{Q}^{\mathrm{T}}) + \beta\mathbf{D}\mathbf{W} = 0, \tag{17}$$

where $\mathbf{D}$ is a diagonal matrix with its diagonal element $\mathbf{D}_{i,i} = 1/(2\|\mathbf{W}_i\|_2 + \tau)$.[3] In this way, the update rule for $\mathbf{W}$ can be formulated as follows:

$$\mathbf{W} \leftarrow [\alpha(\mathbf{P}\mathbf{P}^{\mathrm{T}} + \mathbf{Q}\mathbf{Q}^{\mathrm{T}}) + \beta\mathbf{D}]^{-1}\alpha(\mathbf{P}\mathbf{P}^{\mathrm{T}} + \mathbf{Q}\mathbf{Q}^{\mathrm{T}}). \tag{18}$$

The time complexity of the inverse operation in Eq. (18) is $O(N^3)$. Since matrix $(\mathbf{P}\mathbf{P}^{\mathrm{T}} + \mathbf{Q}\mathbf{Q}^{\mathrm{T}})$ is symmetric, to improve the computational efficiency of the inverse operation, we consider the approximation $(\mathbf{P}\mathbf{P}^{\mathrm{T}} + \mathbf{Q}\mathbf{Q}^{\mathrm{T}}) \approx \mathbf{F}\mathbf{F}^{\mathrm{T}}$, where $\mathbf{F} \in \mathbb{R}^{N \times N'}$, $N' \ll N$. In this way, matrix $\mathbf{F}$ can be obtained by first using the singular value decomposition (SVD) to factorize matrix $(\mathbf{P}\mathbf{P}^{\mathrm{T}} + \mathbf{Q}\mathbf{Q}^{\mathrm{T}})$, and then selecting the singular vectors corresponding to the top N' largest singular values. Based on Woodbury matrix identity, the inverse operation for updating $\mathbf{W}$ can be approximately written as

$$[\alpha(\mathbf{P}\mathbf{P}^{\mathrm{T}} + \mathbf{Q}\mathbf{Q}^{\mathrm{T}}) + \beta\mathbf{D}]^{-1} \approx \frac{1}{\beta}\mathbf{D}^{-1} - \frac{\alpha}{\beta^2}\mathbf{D}^{-1}\mathbf{F}(\mathbf{I} + \frac{\alpha}{\beta}\mathbf{F}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{D}^{-1}. \tag{19}$$

With the operation in Eq. (19), the time complexity of the inverse operation in Eq. (18) can be reduced to $O(N^2N')$. The time complexity for updating $\mathbf{W}$ is $O[N^2(M + J + N')]$. Algorithm 1 describes the overall optimization algorithm of MTMA. At each iteration, it requires $O(KM^2)$ and $O(KJ^2)$ to calculate $\widetilde{\mathcal{R}}^{(1)}_{i,j,k}$ and $\widetilde{\mathcal{R}}^{(2)}_{i,j,k}$, respectively, which is the most time-consuming step in the

[3] Due to the row sparsity of matrix $\mathbf{W}$, $||\mathbf{W}_i||_2$ could be very close to 0. Therefore, we define $\mathbf{D}_{i,i} = 1/(2||\mathbf{W}_i||_2 + \tau)$, where $||\cdot||_2$ is the l_2-norm of a vector and τ is a very small constant.

Algorithm 1 MTMA optimization algorithm

Input: $\mathcal{R}$, $\mathbf{U}$, $\mathbf{V}$, N, N', M, J, K, σ, α, β, L;
Output: $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{W}$, $\mathcal{S}^{(1)}$, $\mathcal{S}^{(2)}$.

```
Initialize matrices P, Q, W, A, B, and C randomly;
Calculate matrices Θ = (I − W)(I − W)^T and D;
Calculate tensors S^(1) and S^(2) by Eq. (11);
repeat
    for each training sample (d_i, d_j, k, R_{i,j,k}) do
        Calculate R̃^(1)_{i,j,k} and R̃^(2)_{i,j,k} by Eq. (3) and Δ_{i,j,k} = R_{i,j,k} − R̃_{i,j,k};
        Update P_i, Q_j, A, B, and C by Eq. (14) and Eq. (15);
        Update tensors S^(1) and S^(2) by Eq. (11);
    end for
    F = SVD(PP^T + QQ^T, N');
    Update matrix W by Eq. (18) and Eq. (19);
    Update matrices Θ and D;
until convergence or the maximal iteration reaches.
```

SGD algorithm. According to Eq. (14), the time complexities for computing the partial derivatives of $\widetilde{\mathcal{R}}_{i,j,k}$ with respect to $\mathbf{P}_i$, $\mathbf{Q}_j\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are $O(LM)$, $O(LJ)$, $O(L^2{+}LM)$, $O(L^2{+}LJ)$, and $O(L^2{+}LM{+}LJ)$, respectively. Due to $L \ll M, J, K$, the total computational cost for updating $\mathbf{P}_i$, $\mathbf{Q}_j\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in Eq. (15) is $O[K(M^2+J^2)]$. Therefore, the time complexity for the optimization algorithm of MTMA is $O[|\mathcal{S}|K(M^2+J^2)+N^2(M+J+N')]$, where $|\mathcal{S}|$ is the number of training samples.

5. Experiments

In this section, we conduct experiments on the real-world dataset to evaluate the performance of MTMA. Here, we first introduce the real-world dataset and experimental settings. Then, MTMA is compared with nine baseline methods and its three variants to evaluate its performance in ADDI prediction. After that, the sparsity study and parameter analysis of MTMA are presented. Finally, we conduct a case study for visualized presentation to further validate the effectiveness of MTMA in exploring the adverse mechanisms and leading factors of ADDIs. All experiments are conducted with Python 3.6 on a PC under Windows 8.1 with Intel (R), 3.30 GHz and 16 GB real memory.[4]

5.1. Dataset

FAERS[5] is a public database that contains adverse event reports submitted to FDA. Mining from FAERS, Tatonetti et al. [25] built TWOSIDES that records the polypharmacy adverse interactions caused by the co-administration of two drugs. There are total 63,473 distinct adverse drug pairs associated with $K = 1318$ adverse interactions in TWOSIDES. In this paper, we extract 576,513 ADDIs from TWOSIDES that are associated with $N = 555$ drugs. The molecular structures of drugs can be obtained from DrugBank [28]. We extract canonical SMILES from DrugBank and use PubChem substructure fingerprints [29] to construct the ground-truth molecular structure matrix $\mathbf{U}$. Each drug is represented by an 881 dimensional binary vector ($M = 881$) such that if drug d_i owns the jth molecular substructure, then $\mathbf{U}_{ij} = 1$; otherwise $\mathbf{U}_{ij} = 0$. The ground-truth side effect matrix $\mathbf{V}$ is established based on SIDER database [58]. Each drug is associated with a 1656 dimensional binary vector ($J = 1656$) such that if drug d_i owns the jth side effect, then $\mathbf{V}_{ij} = 1$; otherwise $\mathbf{V}_{ij} = 0$.

5.2. Experimental settings

We carry out 10-fold cross validation to estimate the performance of different methods. To be specific, we randomly divide the ADDI dataset into 10 parts. In each experiment, 90% ADDIs are sampled and their corresponding molecular structures and side effects are adopted for training, and then the remaining 10% ADDIs are used for testing. For a fair comparison, six metrics, i.e., Accuracy, Precision, Recall, F-score, AUC, and AUPR, are adopted to measure the results of different methods. According to the ground-truth and the results output by the predictive models, there are four possible outcomes: (1) an ADDI is recognized as an ADDI (TP); (2) an ADDI is recognized as a non-ADDI (FN); (3) a non-ADDI is recognized as an ADDI (FP); (4) a non-ADDI is recognized as a non-ADDI (TN). Therefore, we have

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN},$$
$$Precision = \frac{TP}{TP+FP},$$
$$Recall = \frac{TP}{TP+FN},$$
$$F-score = 2 \times \frac{Precision \times Recall}{Precision + Recall}.$$

AUC is the area under the ROC curve, which is a plot of true positive rate ($tpr = \frac{TP}{TP+FN}$) vs. false positive rate ($fpr = \frac{FP}{FP+TN}$). AUPR is the area under the precision–recall curve. The higher the values of the six metrics, the better the prediction performance of a method.

To evaluate the performance of MTMA, we compare it with the following nine baseline methods:

- **MSSA**: MSSA is a large-scale learning model that measures the molecular structure similarities among drugs and combines the knowledge of known interactions for ADDI prediction [14];
- **LPSE**: LPSE is a label propagation ADDI prediction method based on the clinical side effects of drugs [7];
- **CPFS**: CPFS is a computational method that utilizes the functional attribute of drugs and computes their inner product-based similarities to predict ADDIs [40];
- **MDM**: MDM is a multi-task dyadic regression method by employing molecular structure to predict specific types of ADDIs [18];
- **DLP**: DLP is a deep learning method by using molecular structure and drug name to predict adverse drug–drug and drug–food interactions [31];
- **TSSL**: TSSL is a statistical learning model that measures topological and semantic similarities among drugs and uses link prediction method for ADDI prediction [27];
- **MLMA**: MLMA is a machine learning based ADDI prediction method by integrating multiple attribute similarities among drugs [33];
- **SMLS**: SMLS is a similarity-based model that designs an applicable protocol for large-scale ADDI prediction [5];
- **SFLLN**: SFLLN is a sparse feature learning ensemble model with linear neighborhood regularization by introducing drug attribute-based manifold regularization into matrix factorization for ADDI prediction [32].

To further validate the effectiveness of the proposed model, we also compare MTMA with its following three variants for

[4] The source codes of MTMA are available at https://github.com/AdverseDDI/MTMA.git.

[5] http://open.fda.gov/data/faers/

analyzing the effects of the supervision of attributes, the self-representation of drug attributes, and the exploration of leading factors on ADDI prediction.

- **MTMA-S**: MTMA-S is a variant of MTMA that does not consider the supervision of attributes for ADDI prediction. Here, we set matrices **U** and **V** to be **0**;
- **MTMA-R**: MTMA-R is a variant of MTMA that models the adverse interactions among drugs without the consideration of the self-representation of drug attributes. We set parameter α to be 0;
- **MTMA-E**: MTMA-E is another variant of MTMA that does not explore the leading molecular substructures and side effects for each ADDI. In particular, parameter β is set to be 0.

Each comparison algorithm owns different regularization parameters. The parameters for each comparison method are set following the experimental settings in its original work. For a fair comparison, a grid search strategy is adopted to tune the parameters of MTMA and output the best results. Specifically, we use 10-fold cross validation to determine parameters σ, α, and β chosen from {0.001, 0.01, 0.1, 0.5, 1, 5, 10, 100, 500, 1000} and the rank L of tensors $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ set as {2, 4, 6, …, 18, 20}. In addition, we empirically set learning rate $\lambda = 10^{-4}$ and $N' = 0.1N$. In the experiment, it can be found that MTMA achieves the best results when $\sigma = 5$, $\alpha = 0.5$, $\beta = 100$, and $L = 18$. Detailed analysis of parameter settings for MTMA will be presented in Section 5.5.

5.3. Experimental results

The experimental results of the comparison methods are presented in Table 2. As we can see, the values of all metrics provided by MTMA are higher than those by other methods, which demonstrates the effectiveness of MTMA in ADDI prediction. For example, we observe that MTMA outperforms MSSA by 43.24%, LPSE by 19.89%, CPFS by 37.68%, MDM by 22.14%, and DLP by 12.24% in terms of AUC. In essential, ADDIs are generally caused by the interactions between multiple drug attributes and integrating multiple attributes into the same model can improve the prediction results. These five methods only incorporate one attribute for ADDI prediction, which makes it difficult for them to well model the adverse interactions among drugs.

Additionally, although TSSL, MLMA, SMLS, and SPLLN take multiple attributes into consideration, it can be observed that their AUC values are still lower than that output by MTMA. On the one hand, these methods mainly focus on measuring multiple attribute similarities among drugs or mapping the attributes into an integrated interaction space for ADDI prediction, but do not estimate the adverse values contributed from different attributes. On the other hand, the self-representation of drug attributes is an important intrinsic property for ADDI modeling, due to that the attributes of a drug can be self-represented by the attributes of a few representative drugs. These four methods do not incorporate this underlying property into ADDI modeling, accounting for their relatively low prediction performance. In addition, among the comparison methods, only MTMA designs interpretable tensors and incorporates $l_{2,1}$-norm regularization to explore adverse mechanisms and leading factors of ADDIs, which may account for its superior performance over other methods in ADDI prediction.

By comparing MTMA with its variants, we find that the average improvements of MTMA over MTMA-S, MTMA-R, and MTMA-E in terms of AUC are 11.81%, 5.60%, and 2.79%, respectively. First, the ground-truth molecular structure and side effect information of drugs are incorporated as spherical Gaussian priors into MTMA to guide the estimation of the predicted molecular structure and side effect matrices, which finally evolves as the minimum of the Frobenius norm to calculate the matrix differences between the predicted attribute matrices and the ground-truth ones. In this way, the predicted attribute matrices can be more consistent with the ground-truth ones and the interpretable tensors $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ output by MTMA can be better to capture the relationships between attributes and adverse interactions, leading to a better performance of MTMA over MTMA-S. Second, MTMA considers the self-representation of drug attributes by adopting the attribute information of a few representative drugs to reconstruct the predicted attribute matrices. This self-representation regularization not only introduces a coefficient matrix **W** to be regarded as a linear transformation between the attributes of representative drugs and the predicted attribute matrices, but also establishes an underlying relations between molecular structure and side effect, leading to the superior performance of MTMA over MTMA-R. Third, due to that ADDIs are caused by the incompatibility of a few privileged molecular substructures and the aggravation of a few significant side effects, we impose $l_{2,1}$-norm on predicted molecular structure and side effect matrices to formularize this issue, which results in a better performance of MTMA than MTMA-E. According to the experimental results between MTMA and its variants, we can conclude that the supervision of drug attributes, the self-representation of drug attributes, and the exploration of leading factors can benefit ADDI prediction.

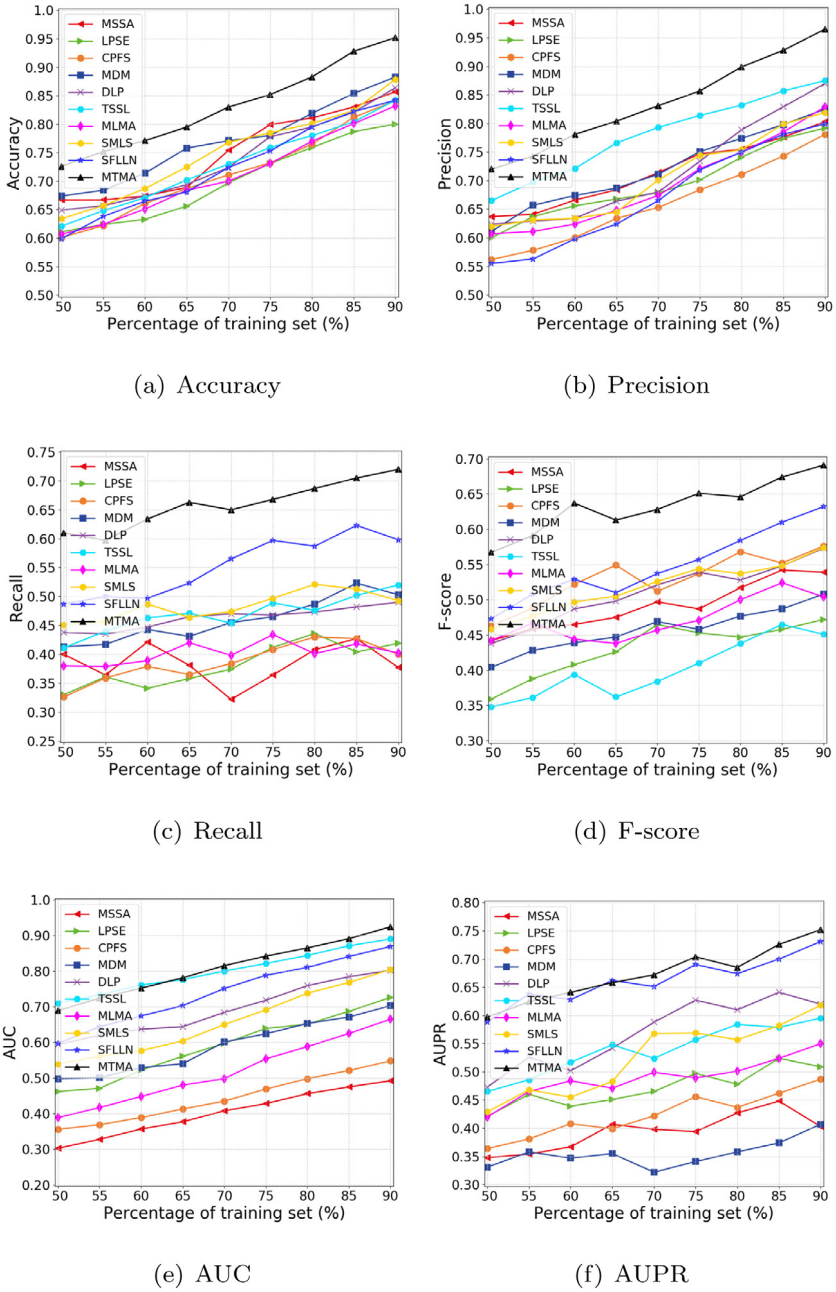
To further validate the performance of MTMA in ADDI prediction, we conduct additional experiments by first sampling {50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%} of ADDIs and their corresponding attribute vectors for training, and then using the remaining samples for testing. Fig. 5 shows the experimental results of the comparison methods in terms of six metrics with the variation of training sets. As we can see from Figs. 5(a), 5(b), and 5(e), the increase of training set correspondingly leads to the enhancements of Accuracy, Precision, and AUC. However, the values of Recall, F-score, and AUPR do not increase along with the training set, as shown in Figs. 5(c), 5(d), and 5(f), respectively. This is possibly because the number of ADDIs is less than the number of non-ADDIs in the experimental dataset, leading to that non-ADDIs are more inclined to be learned by the comparison methods than ADDIs. In this way, ADDIs are likely to be mistakenly predicted as non-ADDIs and the value of FN is relatively high in most cases. Meanwhile, we observe that the value of FN is more sensitive to the variation of training sets than the values of TP, FP, and TN. These reasons may primarily lead to the fluctuation of the curves in Figs. 5(c), 5(d), and 5(f). In addition, it can be seen from Fig. 5(f) that the values of AUPR output by SFLLN in some cases are higher than those output by MTMA. The main reason may lie on that SFLLN considers the sparsity of drug attribute spaces and adopts sparse feature learning to map different drug attribute spaces into the common interaction space. These results inspire us to further investigate the intrinsic properties of drug attribute spaces for improving ADDI prediction. From the results in Fig. 5, we can also conclude that MTMA can achieve better performance than other methods in most cases.

5.4. Sparsity study of MTMA

In MTMA, $l_{2,1}$-norm is imposed on the predicted molecular structure matrix **P** and side effect matrix **Q** to constrain their row sparsity so as to explore the leading molecular substructures and side effects of each specific ADDI. In this subsection, we aim to study the effects of the row sparsity of matrices **P** and **Q** on the performance of MTMA. Here, only the results of the row sparsity of matrix **P** in terms of AUC and AUPR are presented in Fig. 6, as the results of sparsity study for matrix **Q** are similar to those for matrix **P**. It can be seen from Figs. 6(a) and 6(b) that the higher

Table 2
Experimental results of comparison methods for ADDI prediction.

Method	Accuracy	Precision	Recall	F-score	AUC	AUPR
MSSA	0.8571 ± 0.2057	0.8030 ± 0.1039	0.3771 ± 0.2001	0.5392 ± 0.1349	0.4923 ± 0.1544	0.4037 ± 0.1969
LPSE	0.8002 ± 0.0287	0.7923 ± 0.0456	0.4193 ± 0.0257	0.4719 ± 0.2054	0.7258 ± 0.0011	0.5092 ± 0.0032
CPFS	0.8410 ± 0.0329	0.7810 ± 0.2411	0.4004 ± 0.1023	0.5763 ± 0.1410	0.5479 ± 0.3011	0.4871 ± 0.2101
MDM	0.8832 ± 0.0347	0.8249 ± 0.0421	0.5029 ± 0.1294	0.5077 ± 0.1217	0.7033 ± 0.0474	0.4067 ± 0.1483
DLP	0.8645 ± 0.1287	0.8708 ± 0.0746	0.4897 ± 0.0977	0.5734 ± 0.0321	0.8023 ± 0.1049	0.6204 ± 0.0284
TSSL	0.8424 ± 0.1847	0.8757 ± 0.1050	0.5200 ± 0.0849	0.4512 ± 0.0580	0.8900 ± 0.1974	0.5955 ± 0.1082
MLMA	0.8320 ± 0.1094	0.8291 ± 0.1010	0.4023 ± 0.0891	0.5044 ± 0.1017	0.6651 ± 0.0125	0.5497 ± 0.0475
SMLS	0.8793 ± 0.0381	0.8193 ± 0.0312	0.4932 ± 0.0421	0.5745 ± 0.1278	0.8041 ± 0.0187	0.6175 ± 0.1078
SFLLN	0.8424 ± 0.1031	0.7987 ± 0.1020	0.5983 ± 0.1004	0.6322 ± 0.1271	0.8692 ± 0.0154	0.7310 ± 0.0791
MTMA-S	0.8741 ± 0.0198	0.8268 ± 0.0899	0.5758 ± 0.0779	0.5524 ± 0.0871	0.8066 ± 0.0687	0.5811 ± 0.1046
MTMA-R	0.8559 ± 0.0455	0.9007 ± 0.0577	0.5899 ± 0.1004	0.6418 ± 0.1107	0.8687 ± 0.1200	0.6341 ± 0.0944
MTMA-E	0.9004 ± 0.0258	0.8991 ± 0.0484	0.6128 ± 0.0478	0.5987 ± 0.0782	0.8968 ± 0.0448	0.6522 ± 0.0587
MTMA	**0.9516 ± 0.0145**	**0.9651 ± 0.0427**	**0.7202 ± 0.0137**	**0.6911 ± 0.0387**	**0.9247 ± 0.0069**	**0.7515 ± 0.0549**

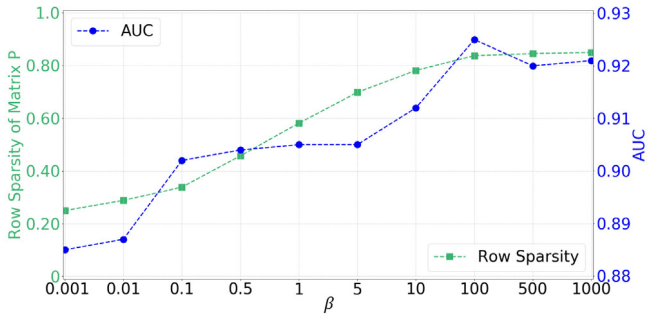


(a) Accuracy (b) Precision

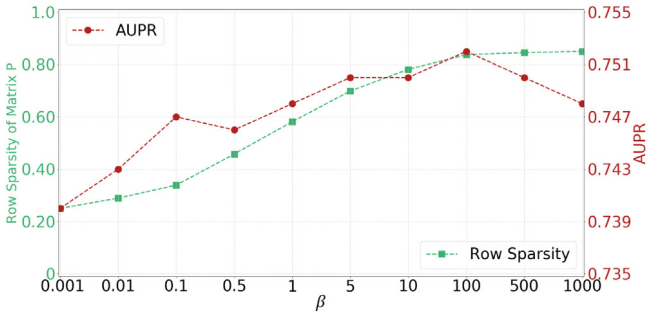
(c) Recall (d) F-score

(e) AUC (f) AUPR

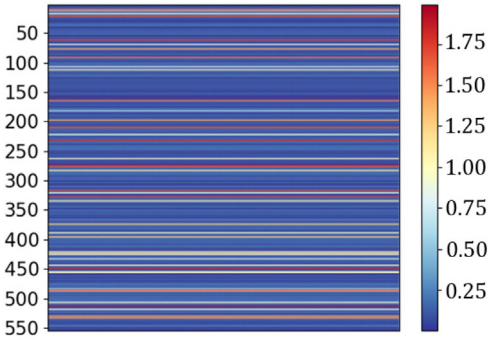
Fig. 5. Experimental results of the comparison algorithms in terms of six metrics with the variation of training sets.



(a) The effects of β on row sparsity and AUC.



(b) The effects of β on row sparsity and AUPR.



(c) The visualization of row sparsity for $\mathbf{P}$.

Fig. 6. The results of row sparsity of matrix $\mathbf{P}$ in terms of AUC and AUPR.

the value of parameter β leads to the better the row sparsity of matrix $\mathbf{P}$, but does not lead to the higher the values of AUC and AUPR. These results indicate that there is a balance between the row sparsity of matrix $\mathbf{P}$ and the performance of MTMA. We find that when $\beta = 100$, MTMA outputs the best results with AUC = 0.9247, AUPR = 0.7515, the row sparsity of matrix $\mathbf{P}$ is 0.8374 (i.e., the l_2-norm of 83.74% rows are close to 0), and its visualization of $l_{2,1}$-norm is shown Fig. 6(c). Each row in Fig. 6(c) represents to the value of l_2-norm for the corresponding row in matrix $\mathbf{P}$. The deeper the blue, the closer the value of l_2-norm is to 0. It can be observed from Fig. 6 that MTMA is capable of estimating matrix $\mathbf{P}$ with good row sparsity.

5.5. Parameter analysis

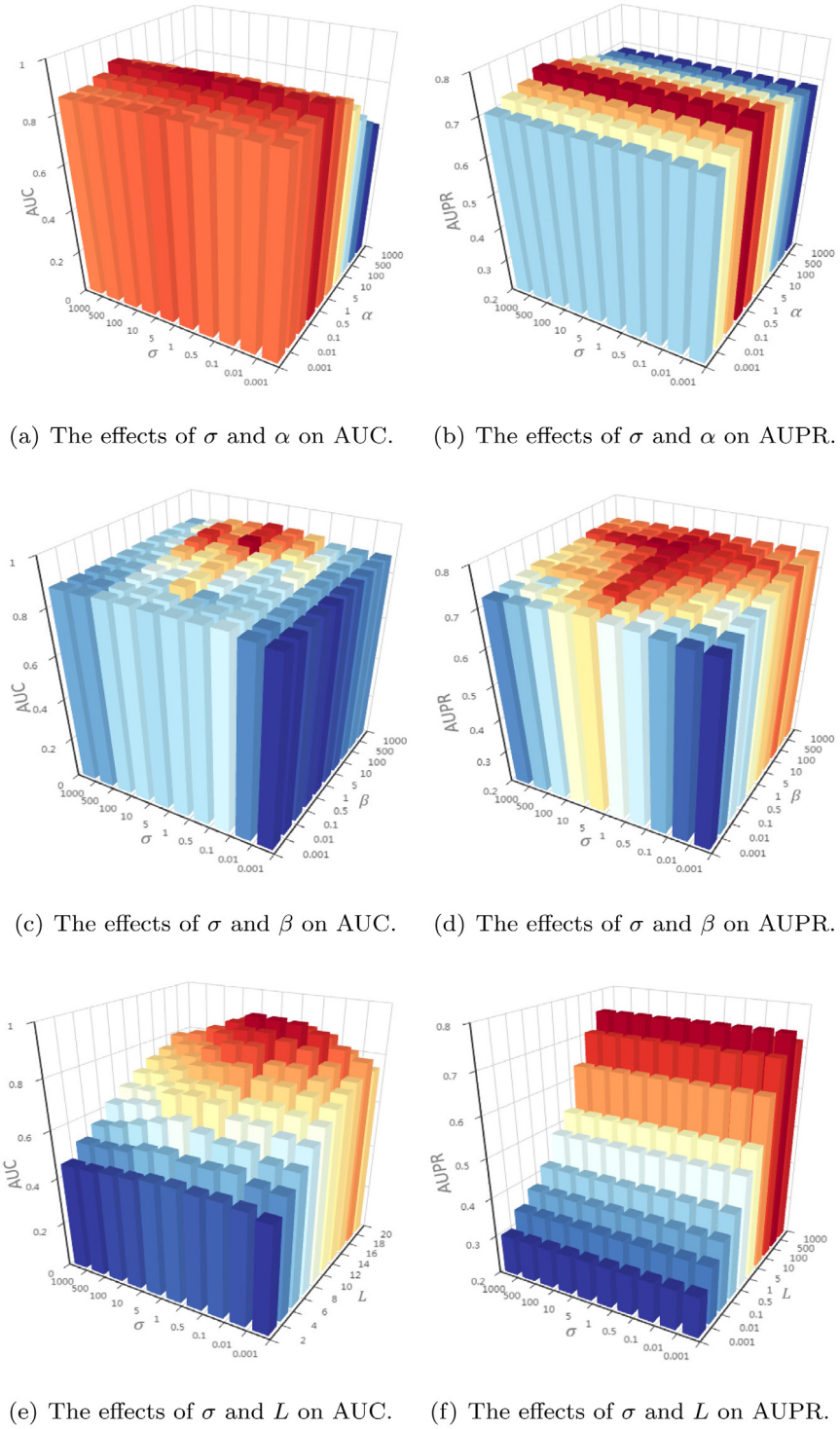
As described in Section 4, MTMA contains four parameters σ, α, β, and L. Based on the grid search strategy, we investigate the effects of each parameter by fixing the others to analyze its impacts on the performance of MTMA. Fig. 7 shows the results of σ, α, β, and L in terms of AUC and AUPR. It can be seen that the variations of σ, α, β, and L have different impacts on the values AUC and AUPR. In particular, we can observe from Figs. 7(a), 7(c), and 7(e) that when σ is fixed, the variation of β has less impacts on the value of AUC than those of α and L, which indicates that the sensitivity of α and L in terms of AUC is higher than β. By comparing Figs. 7(b), 7(d), and 7(f), we find that when σ is fixed, the value of AUPR is more sensitive to the variation of L than those of α and β. From all the results, we can observe that MTMA can obtain the best results when $\sigma = 5$, $\alpha = 0.5$, $\beta = 100$, and $L = 18$.

5.6. Case study

The main advantage of MTMA over the baseline methods is its capability in exploring the adverse mechanisms among drugs and revealing the leading factors of ADDIs. To be specific, we design two interpretable tensors $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ to capture the relationships between attributes and adverse interactions and adopt $l_{2,1}$-norm to constrain the row sparsity of predicted attribute matrices $\mathbf{P}$ and $\mathbf{Q}$ so as to explore the leading molecular substructures and significant side effects of each specific ADDI. These interpretable information can provide valuable guidance for drug safety studies to detect potential ADDIs in laboratory tests and reduce the research expenditure for drug developments.

To illustrate this, we conduct a case study of four examples for visualized presentation, as shown in Fig. 8, in which the molecular structure and some significant side effects for each drug are presented. In particular, the leading molecular substructures that may cause specific adverse interactions are covered by ellipses. It can be seen from Fig. 8(a) that the leading molecular substructures for *Atropine* and *Prednisolone* are covered by red and green ellipses, respectively, which may result in the increase of intraocular pressure. Meanwhile, cardioacceleration, a significant side effect caused by *Atropine*, may aggravate the side effect of cardiomegaly induced by *Prednisolone*, resulting in ventricular fibrillation caused by their interaction. As we can see from Fig. 8(b), the co-administration of *Pravastatin* and *Gemfibrozil* can result in acute renal failure and rhabdomyolysis. In particular, acute renal failure is caused by the incompatibility of the leading molecular substructures of *Pravastatin* and *Gemfibrozil* that are covered by lavender and blue ellipses, respectively, while rhabdomyolysis is induced by the aggravation of arthralgia caused by *Pravastatin* and myasthenia caused by *Gemfibrozil*. In terms of *Aspirin* and *Warfarin* shown in Fig. 8(c), the leading molecular substructures of these two drugs may result in gastrointestinal hemorrhage, which also results from the leading side effects gastrointestinal ulcer and ecchymosis caused by *Aspirin* and *Warfarin*, respectively. In addition, we find that excessive bleeding can be also caused by the aggravation of these two leading side effects. As we have mentioned, the combination of *Zidovudine* and *Acyclovir* can result in nausea, vomiting, renalinsufficiency, and lethargy. In the experiment, we find that renalinsufficiency may be caused by the leading molecular substructures of *Zidovudine* and *Acyclovir*, which are covered by purple and orange ellipses, as shown in Fig. 8(d). On the other hand, the adverse interactions nausea and vomiting may be induced by the aggravation of flatulence and abdominal discomfort by *Zidovudine* and *Acyclovir*, respectively. Additionally, the interaction between them may also worsen their side effects hypoxemia and somnolence, which ultimately leads to lethargy.

Meanwhile, the incompatibility of molecular structures and the aggregation of side effects can be also explored by the designed interpretable tensors $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$, respectively. In terms of adverse molecular structure–molecular structure interaction



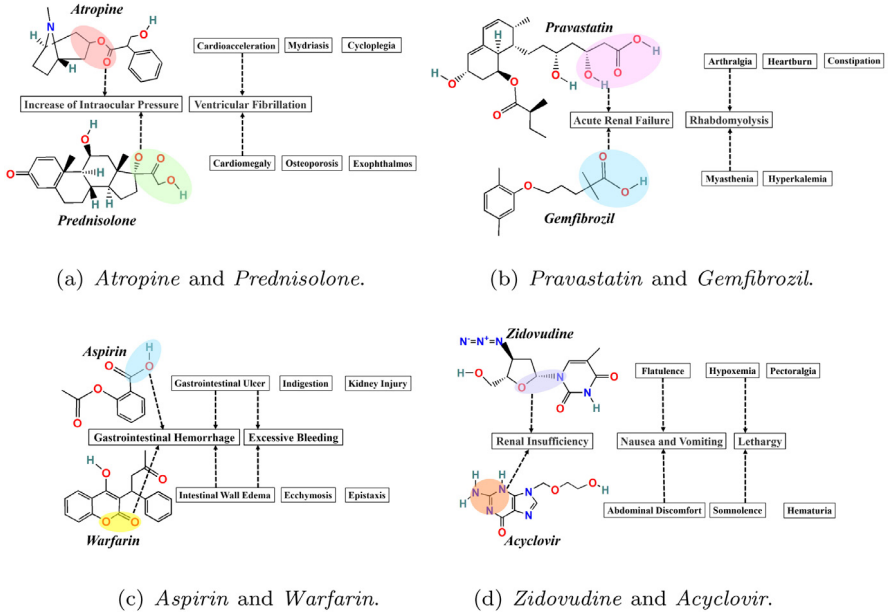
(a) The effects of σ and α on AUC. (b) The effects of σ and α on AUPR.

(c) The effects of σ and β on AUC. (d) The effects of σ and β on AUPR.

(e) The effects of σ and L on AUC. (f) The effects of σ and L on AUPR.

Fig. 7. The effects of parameters σ, α, β, and L on AUC and AUPR.

tensor $\mathcal{S}^{(1)}$, the value of entry $\mathcal{S}^{(1)}_{i,j,k}$ can be seen as an adverse indicator between the ith molecular substructure and the jth molecular substructure with respect to the kth adverse interaction. To be specific, it can be inferred that the interaction between the ith molecular substructure and the jth molecular substructure is likely to result in the kth adverse interaction if the value of $\mathcal{S}^{(1)}_{i,j,k}$ is relatively more significant than many other entries in tensor $\mathcal{S}^{(1)}$. Taking *Aspirin* and *Warfarin* shown in Fig. 8(c) as an example, in the experiments, we find that the adverse interaction gastrointestinal hemorrhage may be caused by their leading molecular substructures covered by blue and yellow eclipses, respectively. Correspondingly, it can be observed from tensor $\mathcal{S}^{(1)}$ that the value of $\mathcal{S}^{(1)}_{i,j,k}$ is higher than many other entries in $\mathcal{S}^{(1)}$, where the ith and the jth molecular substructures are the ones covered by eclipses shown in Fig. 8(c), and the kth adverse interaction corresponds to gastrointestinal hemorrhage. This result indicates the potential incompatibility between the ith and the jth molecular substructures, which may lead to the kth adverse interaction. As a result, it could be inferred that the co-administration of another two drugs may also potentially cause



(a) *Atropine* and *Prednisolone*. (b) *Pravastatin* and *Gemfibrozil*.

(c) *Aspirin* and *Warfarin*. (d) *Zidovudine* and *Acyclovir*.

Fig. 8. The case study of the leading molecular substructures and side effects explored by MTMA.

gastrointestinal hemorrhage if these two drugs respectively own the *i*th and the *j*th molecular substructures shown in Fig. 8(c). Likewise, in adverse side effect-side effect interaction tensor $\mathcal{S}^{(2)}$, we can also infer from $\mathcal{S}^{(2)}_{i,j,k}$ that the *k*th adverse interaction may be caused by the aggregation of the *i*th and the *j*th side effects, which helps to detect the potential adverse interaction if two drugs respectively result in the *i*th and the *j*th side effects. Therefore, these underlying adverse relationships included in tensors $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ can provide significant evidences for predicting the potential adverse interactions among drugs so as to reduce the expenditure in drug developments.

6. Conclusions

In this paper, we propose a novel multi-task multi-attribute (MTMA) learning model for ADDI prediction. In MTMA, two attributes are adopted to model the adverse interactions among drugs and two interpretable tensors are designed to capture the adverse mechanisms among drugs. In addition, $l_{2,1}$-norm is imposed on the predicted attribute matrices for exploring the leading factors of ADDIs. Experimental results on the real-world dataset show that MTMA achieves better performance compared with nine baseline methods and its three variants.

There are still some problems to be solved in the future. First, we find that MTMA is sometimes inferior to the comparison algorithm SFLLN in terms of AUPR. In the future, the intrinsic properties (e.g., orthogonality, low-rank, and positive semi-definite, etc.) of the predicted attribute matrices will be explored to improve the prediction results. Second, we plan to design more complicated integration strategies to learn the contributions of different attributes on ADDI modeling. Third, in practice, the biochemical activities of drugs are usually highly dependent on the molecular structure of drugs [59], which becomes an essential factor that are related to many other attributes. Thus, we aim to incorporate this issue to explore the underlying dependences among attributes. In addition, molecular structure and side effect adopted in this paper are not the only attributes that lead to the adverse interactions among drugs. In the future, more attributes, such as target, enzyme, and pathway will be taken into consideration for further ADDI prediction.

CRediT authorship contribution statement

Jiajing Zhu: Conceptualization, Methodology, Writing - review & editing. **Yongguo Liu:** Writing - review & editing, Supervision, Funding acquisition. **Chuanbiao Wen:** Methodology, Resources, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported in part by the National Key R&D Program of China under Grant 2017YFC1703905 and in part by the Sichuan Science and Technology Program under Grant 2018SZ0065, Grant 2019YFS0019, and Grant 2018TJPT0039.

References

[1] C.O. Plumpton, D. Roberts, M. Pirmohamed, D.A. Hughes, A systematic review of economic evaluations of pharmacogenetic testing for prevention of adverse drug reactions, Pharmacoeconomics 34 (8) (2016) 771–793.
[2] M. Liu, Y. Wu, Y. Chen, J. Sun, Z. Zhao, X. Chen, M.E. Matheny, H. Xu, Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs, J. Am. Med. Inform. Assoc. 19 (E1) (2012) E28–E35.
[3] X. Qin, T. Kakar, S. Wunnava, E.A. Rundensteiner, L. Cao, MARAS: Signaling multi-drug adverse reactions, in: Proceedings of the Twenty-Third ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 1615–1623.
[4] R. Cai, M. Liu, Y. Hu, B.L. Melton, M.E. Matheny, H. Xu, L. Duan, L.R. Waitman, Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports, Artif. Intell. Med. 76 (2017) 7–15.
[5] S. Vilar, E. Uriarte, L. Santana, T. Lorberbaum, G. Hripcsak, C. Friedman, N.P. Tatonetti, Similarity-based modeling in large-scale prediction of drug-drug interactions, Nat. Protoc. 9 (9) (2014) 2147–2163.
[6] X. Chu, Y. Lin, J. Gao, J. Wang, L. Wang, Multi-label robust factorization autoencoder and its application in predicting drug-drug interactions, 2018, pp. 1–9, arXiv preprint arXiv:1811.00208.
[7] P. Zhang, F. Wang, J. Hu, R. Sorrentino, Label propagation prediction of drug-drug interactions based on clinical side effects, Sci. Rep. 5 (2015) 1–10.

[8] N.P. Tatonetti, G.H. Fernald, R.B. Altman, A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports, J. Am. Med. Inform. Assoc. 19 (1) (2012) 79–85.

[9] M. Takarabe, D. Shigemizu, M. Kotera, S. Goto, M. Kanehisa, Network-based analysis and characterization of adverse drug-drug interactions., J. Chem. Inf. Model. 51 (11) (2011) 2977–2985.

[10] M. Herrero-Zazo, I. Segura-Bedmar, P. Martínez, Conceptual models of drug-drug interactions: A summary of recent efforts, Knowl.-Based Syst. 114 (2016) 99–107.

[11] V. Henrike, S. Noel, H. Ruili, J. Tim, F. Darren, A. Natalia, S. Min, I. James, C.P. Austin, D.G. Lloyd, Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries, Nature Biotechnol. 27 (11) (2009) 1050–1055.

[12] J. König, F. Müller, M.F. Fromm, Transporters and drug-drug interactions: Important determinants of drug disposition and effects, Toxicol. Lett. 238 (2) (2015) S49.

[13] B. Percha, R.B. Altman, Informatics confronts drug-drug interactions, Trends Pharmacol. Sci. 34 (3) (2013) 178–184.

[14] S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabadan, C. Friedman, Drug-drug interaction through molecular structure similarity analysis, J. Am. Med. Inform. Assoc. 19 (6) (2012) 1066–1074.

[15] H. Yu, K. Mao, J. Shi, H. Huang, Z. Chen, K. Dong, S. Yiu, Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization, BMC Syst. Biol. 12 (1) (2018) 101–110.

[16] O. R Scott, R.L. Walsky, V. Karthik, E.A. Gaman, H. J Brian, L.M. Tremaine, The utility of in vitro cytochrome P450 inhibition data in the prediction of drug-drug interactions, J. Pharmacol. Exp. Ther. 316 (1) (2006) 336–348.

[17] L.C. Wienkers, T.G. Heath, Predicting in vivo drug interactions from in vitro drug discovery data, Nat. Rev. Drug Discov. 4 (10) (2005) 825–833.

[18] B. Jin, H. Yang, C. Xiao, P. Zhang, X. Wei, F. Wang, Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 2017, pp. 1367–1373.

[19] C. Xue, X. Zhang, W. Cai, Prediction of drug-drug interactions with bupropion and its metabolites as CYP2D6 inhibitors using a physiologically-based pharmacokinetic model, Pharmaceutics 10 (1) (2017) 1–20.

[20] J.M. Siller-Matula, D. Trenk, S. Krahenbuhl, A.D. Michelson, G. Delle-Karth, Clinical implications of drug-drug interactions with P2Y12 receptor inhibitors, J. Thromb. Haemost. 12 (1) (2014) 2–13.

[21] Q.C. Bui, P.M. Sloot, E.M. Mulligen, J.A. Kors, A novel feature-based approach to extract drug-drug interactions from biomedical text, Bioinformatics 30 (23) (2014) 3365–3371.

[22] D. Sridhar, S. Fakhraei, L. Getoor, A probabilistic approach for collective similarity-based drug-drug interaction prediction, Bioinformatics 32 (20) (2016) 3175–3182.

[23] G. Jiang, H. Liu, H.R. Solbrig, C.G. Chute, Mining severe drug-drug interaction adverse events using Semantic Web technologies: A case study, Biodata Min. 8 (1) (2015) 1–12.

[24] J.D. Duke, H. Xu, W. Zhiping, S. Abhinita, S.D. Karnik, L. Xiaochun, S.D. Hall, J. Yan, C. J Thomas, M.J. Overhage, Literature-based drug interaction prediction with clinical assessment using electronic medical records: Novel myopathy associated drug interactions, PLoS Comput. Biol. 8 (8) (2012) 1–13.

[25] N.P. Tatonetti, P.P. Ye, D. Roxana, R.B. Altman, Data-driven prediction of drug effects and interactions, Sci. Transl. Med. 4 (125) (2012) 1–26.

[26] Y. Shen, K. Yuan, M. Yang, B. Tang, Y. Li, N. Du, K. Lei, KMR: Knowledge-oriented medicine representation learning for drug-drug interaction and similarity computation, J. Cheminformatics 11 (2019) 1–16.

[27] A. Kastrin, P. Ferk, B. Leskošek, Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning, PLoS One 13 (5) (2018) 1–23.

[28] D.S. Wishart, C. Knox, A.C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, M. Hassanali, DrugBank: A knowledgebase for drugs, drug actions and drug targets, Nucleic Acids Res. 36 (2008) 901–906.

[29] Y. Wang, J. Xiao, T.O. Suzek, J. Zhang, J. Wang, S.H. Bryant, PubChem: A public information system for analyzing bioactivities of small molecules, Nucleic Acids Res. 37 (2009) 623–633.

[30] M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000) 27–30.

[31] J.Y. Ryu, H.U. Kim, S.Y. Lee, Deep learning improves prediction of drug-drug and drug-food interactions, Proc. Natl. Acad. Sci. USA 115 (18) (2018) E4304–E4311.

[32] W. Zhang, K. Jing, F. Huang, Y. Chen, B. Li, J. Li, J. Gong, SFLLN: A sparse feature learning ensemble method with linear neighborhood regularization for predicting drug-drug interactions, Inform. Sci. 497 (2019) 189–201.

[33] F. Cheng, Z. Zhao, Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties, J. Am. Med. Inform. Assoc. 21 (E2) (2014) E278–E286.

[34] T. Takeda, M. Hao, T. Cheng, S.H. Bryant, Y. Wang, Predicting drug-drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge, J. Cheminformatics 9 (1) (2017) 1–9.

[35] J. Shi, K. Mao, H. Yu, S. Yiu, Detecting drug communities and predicting comprehensive drug-drug interactions via balance regularized semi-nonnegative matrix factorization, J. Cheminformatics 11 (2019) 1–16.

[36] W. Zhang, Y. Chen, D. Li, X. Yue, Manifold regularized matrix factorization for drug-drug interaction prediction, J. Biomed. Inform. 88 (2018) 90–97.

[37] J. Shi, H. Huang, J. Li, P. Lei, Y. Zhang, K. Dong, S. Yiu, TMFUF: A triple matrix factorization-based unified framework for predicting comprehensive drug-drug interactions of new drugs, BMC Bioinform. 19 (14) (2018) 27–37.

[38] M. Shamsipur, S.M. Pourmortazavi, A.A.M. Beigi, R. Heydari, M. Khatibi, Thermal stability and decomposition kinetic studies of Acyclovir and Zidovudine drug compounds, AAPS PharmSciTech 14 (1) (2013) 287–293.

[39] S.D. Brown, M.G. Bartlett, C.A. White, Pharmacokinetics of intravenous Acyclovir, Zidovudine, and Acyclovir-Zidovudine in pregnant rats, Antimicrob. Agents Chemother. 47 (3) (2003) 991–996.

[40] R. Ferdousi, R. Safdari, Y. Omidi, Computational prediction of drug-drug interactions based on drugs functional similarities, J. Biomed. Inform. 70 (2017) 54–64.

[41] J. Klekota, F.P. Roth, Chemical substructures that enrich for biological activity, Bioinformatics 24 (21) (2008) 2518–2525.

[42] C. Phanus-umporn, W. Shoombuatong, V. Prachayasittikul, N. Anuwongcharoen, C. Nantasenamat, Privileged substructures for anti-sickling activity via cheminformatic analysis, RSC Adv. 8 (11) (2018) 5920–5935.

[43] L. Gaboriau, M. Auffret, J. Bene, H. Cattoir-Vue, S. Gautier, Pruritus: A significant side effect with fampridine, Fundam. Clin. Pharmacol. 30 (S1) (2016) 68.

[44] Z. Tolkien, L. Stecher, A.P. Mander, D.I. Pereira, J.J. Powell, Ferrous sulfate supplementation causes significant gastrointestinal side-effects in adults: A systematic review and meta-analysis, PLoS One 10 (2) (2015) 1–20.

[45] A. Stegeman, N.D. Sidiropoulos, On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition, J. Causal Inference 420 (2) (2015) 540–552.

[46] S. Amari, Backpropagation and stochastic gradient descent method, Neurocomputing 5 (4) (1993) 185–196.

[47] Z. Zhao, Z. Yang, L. Luo, H. Lin, J. Wang, Drug-drug interaction extraction from biomedical literature using syntax convolutional neural network, Bioinformatics 32 (22) (2016) 3444–3453.

[48] A. Cocos, A.G. Fiks, A.J. Masino, Deep learning for pharmacovigilance: Recurrent neural network architectures for labeling adverse drug reactions in Twitter posts, J. Am. Med. Inform. Assoc. 24 (4) (2017) 813–821.

[49] D. Huang, Z. Jiang, L. Zou, L. Li, Drug-drug interaction extraction from biomedical literature using support vector machine and long short term memory networks, Inform. Sci. 415 (2017) 100–109.

[50] Y. Huang, T. Li, C. Luo, H. Fujita, S.-J. Horng, Dynamic fusion of multi-source interval-valued data by fuzzy granulation, IEEE Trans. Fuzzy Syst. 26 (6) (2018) 3403–3417.

[51] H. Wang, Y. Yang, B. Liu, H. Fujita, A study of graph-based system for multi-view clustering, Knowl.-Based Syst. 63 (2019) 1009–1019.

[52] Y. Zhang, Y. Yang, T. Li, H. Fujita, A multitask multiview clustering algorithm in heterogeneous situations based on LLE and LE, Knowl.-Based Syst. 163 (2019) 776–786.

[53] L. Zhou, Q. Wang, H. Fujita, One versus one multi-class classification fusion using optimizing decision directed acyclic graph for predicting listing status of companies, Inf. Fusion 36 (2017) 80–89.

[54] X. Chu, Y. Lin, Y. Wang, L. Wang, J. Wang, J. Gao, MLRDA: A Multi-Task Semi-Supervised Learning Framework for drug-drug interaction Prediction, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, 2019, pp. 4518–4524.

[55] A. Covas, S. Diaz-Insa, D. Ezpeleta, J.C. Garcia-Monco, V. Mateos, M. Sanchez del Rio, Relation between different attributes of medication overuse headache and drug groups, in: Proceedings of 13th Congress of the International Headache Society, 2007, pp. 676–677.

[56] J. Ye, Y. Shi, An effective method to detect seam carving, Knowl.-Based Syst. 35 (2017) 13–22.

[57] D. Huang, X. Cai, C. Wang, Unsupervised feature selection with multi-subspace randomization and collaboration, Knowl.-Based Syst. 182 (2019) 1–15.

[58] M. Kuhn, I. Letunic, L.J. Jensen, P. Bork, The SIDER database of drugs and side effects, Nucleic Acids Res. 44 (2016) 1075–1079.

[59] M. Tian, H. Tan, H. Li, C. You, Molecular weight dependence of structure and properties of chitosan oligomers, RSC Adv. 5 (85) (2015) 69445–69452.