Predicting Drug-Target Interactions by Measuring Confidence with Consistent Causal Neighborhood Interventions

Wenting Ye[1], Chen Li[2], Wen Zhang[1,3,4], Jiuyong Li[5], Lin Liu[5], Debo Cheng[5]*, Zaiwen Feng[1,3,4,6,7]*

[1]*College of Informatics, Huazhong Agricultural University, Wuhan 430070, China*
[2]*Graduate School of Informatic, Nagoya University, Chikusa, Nagoya, 464-8602, Japan*
[3]*Hubei Key Laboratory of Agricultural Bioinformatics, Huazhong Agric, Wuhan 430070, China*
[4]*Key Laboratory of Smart Farming for Agricultural Animals, Huazhong Agric, Wuhan 430070, China*
[5]*UniSA STEM, University of South Australia, Adelaide, 5095, Australia*
[6]*Hubei Hongshan Laboratory, Huazhong Agric, Huazhong Agricultural University, Wuhan 430070, China*
[7]*National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China*
**Corresponding authors. chengd@unisa.edu.au and Zaiwen.Feng@mail.hzau.edu.cn*

Abstract

Predicting drug-target interactions (DTI) is a crucial stage in drug discovery and development. Understanding the interaction between drugs and targets is essential for pinpointing the specific relationship between drug molecules and targets, akin to solving a link prediction problem using information technology. While knowledge graph (KG) and knowledge graph embedding (KGE) methods have been rapid advancements and demonstrated impressive performance in drug discovery, they often lack authenticity and accuracy in identifying DTI. This leads to increased misjudgment rates and reduced efficiency in drug development. To address these challenges, our focus lies in refining the accuracy of DTI prediction models through KGE, with a specific emphasis on causal intervention confidence measures (CI). These measures aim to assess triplet scores, enhancing the precision of the predictions. Comparative experiments conducted on three datasets and utilizing 9 KGE models reveal that our proposed confidence measure approach via causal intervention, significantly improves the accuracy of DTI link prediction compared to traditional approaches. Furthermore, our experimental analysis delves deeper into the embedding of intervention values, offering valuable insights for guiding the design and development of subsequent drug development experiments. As a result, our predicted outcomes serve as valuable guidance in the pursuit of more efficient drug development processes.

Keywords: Drug-target prediction, knowledge graph, causal intervention, confidence measurement, probability calibration

1. Introduction

Drug-target research is an important research direction in the area of drug research and development, and the prediction of DTI [1] is consequently of paramount importance. The purpose of the drug discovery project is to meet the current clinical needs of the lack of effective medical products [2]. With the emergence of new diseases and the development of drug resistance, the demand for new drugs is also increasing. The traditional drug development process may take 12-15 years and cost more than USD 1 billion [3], and the success rate of research and development cannot be guaranteed. The process of bringing a new drug to market consists of four main stages: drug discovery, preclinical testing, clinical trials, and Food and Drug Administration (FDA) review [4]. The core of the drug discovery stage is to obtain disease-related biological information, determine drug-targets, and determine DTI.

Computational prediction of DTI has become an important task in the process of drug discovery [1]. By promoting an understanding of drug action mechanisms, disease pathology, and potential side effects [5, 6, 7], computational prediction of DTI enables the creation of more precise drug-targets. Additionally, it enhances the success rate of novel drug research and development endeavors [8, 9], leading to shortened timelines and decreased development costs. However, with the aid of computer technology and bioinformatics, drug target discovery has increasingly turned to information technology. Sequence and structure alignments [10, 11], homology modeling [12], and deep learning-

based methods [13], which have been widely applied for the drug-target discovery, such as support vector machine, random forest, deep neural network, KGE and other research methods. They are used to mine the feature information of drugs and proteins from a large number of biomedical data.

The research direction of this paper is the DTI represented by KGE. By adding the existing information on drugs and proteins to the KG, the corresponding embedding vector can be generated by using the KGE, to represent the relationship between drugs and proteins and realize the link prediction of drug-targets. However in recent years, many models have emerged in the field of DTI, and breakthroughs and effects have been achieved. However, previous studies have found that existing confidence measurement methods present numerous issues. For instance, the use of the original confidence score or the ranking based on the model output's confidence score renders the output uninterpretable. Additionally, rank-based confidence measurement methods often overlook differences between scores, and their stability cannot be guaranteed.

To tackle the aforementioned challenges, we implement the confidence measure technique of causal intervention for link prediction in drug-target discovery. Initially, the method utilizes the concept of causal intervention to modify the embedding representation. Subsequently, the KGE model is employed to reconstruct a new triplet consisting of drug intervention entities for re-scoring. Finally, a new confidence score is derived through consistency calculation. Unlike the prior study that solely emphasized optimal scoring, this approach evaluates the robustness of the embedded representation results by actively intervening in the input of the entity vector. Our goal is to precisely represent the confidence probability of the model's predictions and enhance the performance of the KGE model in DTI tasks.

The work [14] shows that the causal intervention approach can improve the model's confidence score, but it has not been applied in the area of drug discovery. In this study, we combined the probability calibration method with the causal intervention confidence measurement method to enhance the authenticity of DTI prediction results. The effectiveness of this approach was verified by comparing it with two traditional confidence measurement methods. This was done to address the issue of insufficient model confidence in drug-target predictions based on the KGE model, aiming to improve both the accuracy and stability of these predictions.

Following the work of [15], this study tackles the identified issue and expands upon it by integrating a wider array of models and datasets. Furthermore, it provides a more detailed analysis of the algorithm's functionality and performance. The main contributions are summarized as follows:

- Introduction of three new KGE models: To extend the comparison of the performance of different models and verify the robustness of previous experimental results, we introduce three new KGE models, named TransR, HolE, and TuckER.
- Enhanced method section: A comprehensive introduction to causality is provided, complemented by intricate formulas and pseudo-code, which are employed to elucidate our algorithm.
- Extensive experiments were conducted: The incorporation of a dataset serves to enhance the scope of the experiment. Additionally, extensive experiments were introduced to analyze hyperparameter thresholds and scrutinize intervention substitution values.

The remaining sections are structured as follows: Section 2 provides a comprehensive review of related literature in the KGE, drug-target link prediction and confidence methods. Section 3 overviews the methodology employed in this study. In Section 4, we present the experimental setup and results of our analysis. Finally, Section 5 concludes the paper with a summary of key findings and suggestions for future research directions.

2. Related Work

In this section, we mainly survey some relevant research progress and technical methods in drug link prediction, as well as the prior knowledge of $e - r$ query and confidence methods involved in this paper.

2.1. Knowledge Graph Embeddings

KGE is a technique that represents entities and relationships in a KG as continuous vectors. It aims to evaluate triples in a KG by learning a scoring function. In this process, entities and relations are mapped into a low-dimensional vector space, so that their semantic information is preserved. Let E and R represent the sets of entities and relations, respectively. A KG G comprises factual triples (h, r, t), where $h, t \in E$ and $r \in R$. The cardinalities $|E|$ and $|R|$ indicate the number of entities and relations in G, respectively. The objective of KGE is to depict each entity $e \in E$ (or relation $r \in R$) as a continuous vector in d-dimensional space. Simultaneously, it aims to learn a scoring function f: $E \times R \times E \rightarrow R$ for

evaluating each triple. In general, the KGE model optimizes a loss function and trains the model by minimizing the difference between positive and negative examples. In this process, the model attempts to improve the correct triple score and reduce the wrong triple score, thereby improving the accuracy of the new entity prediction.

In the $e - r$ query, the standard link prediction task aims to determine the target entity $m \in E$ from the entity set E, so that there are triples (e, r, m) or (m, r, e) in the KG G. The key to this task is that a KGE model is needed to evaluate all candidate triples and generate a sequence sorted by scores. In this sort sequence, the entity with the highest score will be specified as m, forming a new triple. This process is based on the semantic association between entities and relationships learned by the model, and the relative importance between triples to ensure that the generated triples have the highest semantic coherence and credibility. In this way, the standard link prediction task can effectively supplement the missing information in the KG and improve the integrity and application value of the KG.

In this paper, we have employed 9 KGE models, namely, TransE[16], ComplEx[17], DistMult[18], RotatE[19], TransH[20], CrossE[21], TransR[22], HolE[23], and TuckER[24]. TransE, TransH, and TransR commonly operate within the real-valued space for representation and computation. Specifically, TransH and TransR improve upon the relationship handling capabilities of TransE by introducing hyperplanes or relation-specific spaces. DistMult utilizes a symmetric bilinear model, simplifying the treatment of relational symmetry. Generally, more sophisticated models such as ComplEx and RotatE tend to exhibit superior performance in tasks like link prediction. Table 1 presents the scoring functions of the 9 KGE models utilized in this study, where $\circ$ denotes the element-wise Hadmard product, $\times$ denotes the tensor product, e^c denotes a complex-number vector. D_{hyp} is the hyperbolic distance, $\bigoplus$ is Möbius addition operation. $\langle , \rangle$ represents the dot product operation, and $Re()$ represents taking the real part.

Table 1 Scoring functions of 9 KGE models

Model	Score Function
TransE	$\|\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|\|$
ComplEx	$Re(\mathbf{h}^{cT} diag(\mathrm{M_r^c})\mathrm{t^c})$
DistMult	$\mathbf{h}^T diag(\mathbf{M_r})\mathbf{t}$
RotatE	$\|\|\mathbf{h} \circ \mathbf{r^c} - \mathbf{t}\|\|$
TransH	$D_{hyp}(h \bigoplus r, t)$
CrossE	$\sigma(CrossNet(\mathbf{h}, \mathbf{r}, \mathbf{t}))$
TransR	$\|\|\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp\|\|_2^2$
HolE	$Re(\langle circ(\mathbf{h}, \mathbf{r}, \mathbf{t})\rangle)$
TuckER	$\mathbf{W} \times \mathbf{h} \times \mathbf{r} \times \mathbf{t}$

2.2. Drug-Target Prediction

The importance of drug-target prediction has become increasingly prominent. The wide application of modern advanced technologies such as machine learning, deep learning and knowledge mapping has further promoted the research progress in this field. [25] predicted the interaction mode and intensity of drugs and proteins through molecular simulation technology [26], providing important information for drug-target development. [27] discussed the application and performance of different KGE models in drug discovery, rather than just focusing on an optimal embedding model. They studied in depth how changes in training settings, selection of hyperparameters, initialization of model parameters, and different splits of data sets affect performance. By studying different KGE models, including TransE, RotatE , and DistMult, the authors compared their performance in tasks such as drug-target prediction, drug-side effect prediction, and drug-drug interaction prediction. The research results of this paper show that the KGE model has a good application prospect in drug discovery, but its performance is affected by many factors, which need further optimization and improvement. Among them, all drug-target pairs are sorted from high to low according to the prediction score based on the traditional confidence measurement method.

In computer technology-based classification tasks, the predictive models exhibit inherent uncertainties, making it challenging to accurately assess the credibility of the model outputs. For instance, [28, 29] highlight issues related to poor model interpretability. Researchers in related fields have undertaken efforts to address these challenges. [30] integrated information from the chemical and genomic spaces, [31] explored the chemical properties, biological pathways, phenotypic effects, and network interactions of binding drugs, and [32] developed and integrated characteristic representations of drugs and diseases at various convolutional layers. In addition to drug-target prediction, there are other association prediction methods, such as drug-drug interaction prediction [33]. Recently, [34] proposed a method called drug-microbe-disease (DMD) association prediction, which has achieved satisfactory results in association prediction. [35] using L3N principles to predict Protein-Protein Interactions (PPIs) has achieved satisfactory results. [36] proposed a network enhancement method to identify and weaken spurious

links among drugs, aiming to improve the accuracy of Drug-Drug Interaction (DDI) predictions. In the future, it can also be explored to apply related methodologies to enhance both the predictive accuracy and interpretability of DTI predictions.

To bolster the reliability and stability of these models, it's crucial to measure the confidence of their outputs as a metric for evaluating prediction quality.

2.3. Confidence Measurement Methods

There exist various traditional methods for measuring confidence, such as probability-based techniques. One common approach involves calculating the model's confidence by assessing the error between the predicted value and the actual value [37]. Typical indicators used include mean square error, as well as mean absolute error. Additionally, statistical methods rooted in statistical knowledge and theory are utilized for confidence measurement, with commonly used indicators encompassing hypothesis testing, as well as confidence intervals. In link prediction tasks, two frequently employed traditional confidence measures are the SigmoidMax (SIG) and TopKSoftmax (TOP) confidence measures [38, 39].

Among them, the SIG uses the Sigmoid function to normalize the original sorting score and map it to a range between 0 and 1. Specifically, for a given ranking score sequence S, the SIG converts each score through the Sigmoid function, and the obtained value represents the confidence of the corresponding entity or relationship, which in turn reflects the degree of correlation between it and the query entity or relationship. Although the optimization effect is relatively stable, it has strong saturation.

On the other hand, the TOP evaluates the confidence of the sorted score sequence S based on the strategy of the Softmax function. Different from the SIG, the TOP does not normalize all the sorting scores, but selects the first K scores and then normalizes them by the Softmax function. The advantage of this method is that it can focus on higher-ranking entities or relationships, and put more attention and weight on possible more relevant prediction results, thereby improving the accuracy and reliability of confidence assessment.

Given the sorted score sequence S of a prediction, the two methods measure the confidence score by:

$$\mathrm{P}_{\mathrm{SIG}}(\boldsymbol{S}) = 1/(1 + \mathrm{e}^{-\max^{(S)}}) \tag{1}$$

$$\mathrm{P}_{\mathrm{TOP}}(\boldsymbol{S}) = \max(\mathrm{e}^{\boldsymbol{S}[:K]}/(\sum_{i=1}^{K} \mathrm{e}^{S_i})) \tag{2}$$

Both of them are computed by the single sequence S and focus on the optimal score. Compared with only focusing on the optimal score, these two methods will lead to some errors in the model. To tackle the challenge that rank-based confidence measures make confidence scores lack stability and interpretability, we propose causal intervention for confidence measures, which aims to improve the robustness and interpretability of confidence measures.

3. The Proposed Causal Intervention Method

In this section, we delve into the foundational theory discussed above, elucidating a meticulous approach to measuring confidence in drug-target link prediction through causal intervention. The content primarily encompasses the algorithm's overarching framework, providing a comprehensive breakdown of each component and the holistic process of the entire prediction task. Building upon previous conference paper findings, we have enriched each method's intricacies and provided a deeper exploration of causal intervention, along with the rationale behind the method's motivation.

3.1. Overview

The fundamental concept underlying this research method involves intervening in the dimension value of a drug entity embedding vector. Various strategies can be employed for intervention. Subsequently, the robustness of the score is then utilized as a criterion to gauge the authenticity of the original triple, determined by comparing the consistency of the triple score sequence derived from the original drug entity with that of the intervention entity.

The pivotal steps and processes of the proposed causal intervention confidence measurement model for DTI are elucidated in Fig. 1. Initially, the data's target entities are encoded into the KGE model for representation learning. Subsequently, the well-learned embedding model is utilized to forecast the links between drugs, targets, and their relationships within the dataset. Each triplet is assigned a score, forming a scoring sequence. The intervention then occurs in the embedding representation of drug entities, facilitating causal intervention confidence measurement. Finally, the results are calibrated by two common probability calibration methods. The performance of these confidence measurement methods is scrutinized and compared.

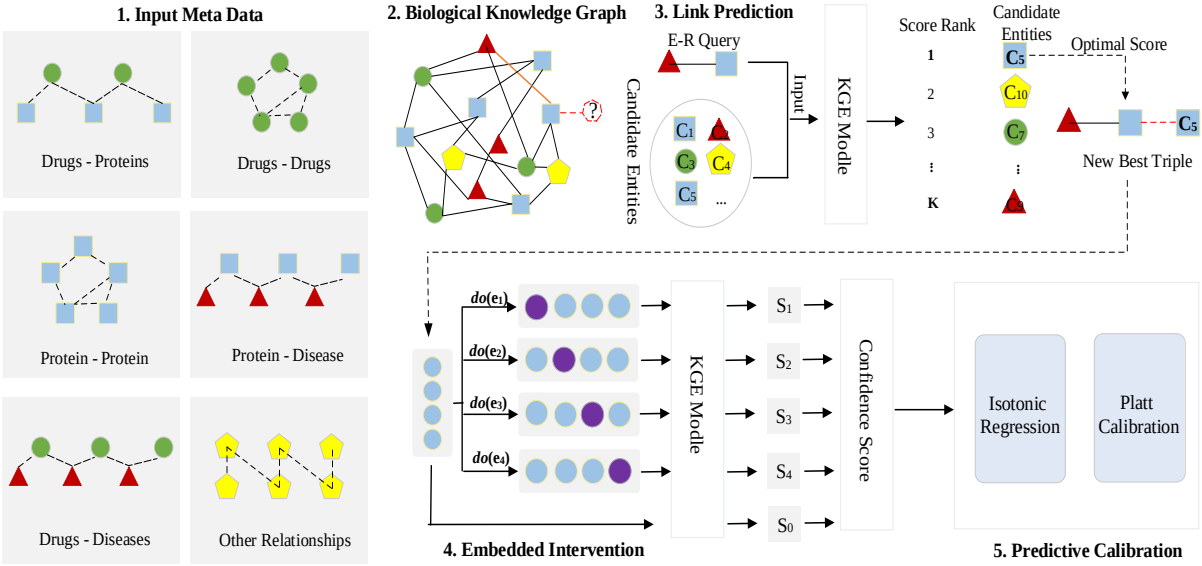

Fig. 1. The architecture illustrates the proposed causal intervention method, which consists of 5 steps: establishing biological knowledge map with metadata, link prediction, neighborhood intervention, consistency score calculation and probability calibration. The link prediction is to obtain the first K candidate entities. And further images are provided later to show details of other steps.

Algorithm 1 Causal Intervention Confidence Algorithm for Confidence Measures

Require: entity $\boldsymbol{e}$, replacement value V, the number of candidates K, pseudo-entity $\boldsymbol{P_e}$, The original scoring sequence $\boldsymbol{S}$

Ensure: Operation 1 affects only one dimension of e

$\boldsymbol{P_e} \Leftarrow$ EmbedAndIntervene($\boldsymbol{e}, V$)
$\boldsymbol{S^P} \Leftarrow$ Sort(CalScores($\boldsymbol{P_e} \bigwedge \boldsymbol{e}$))
Initialize array $\boldsymbol{ConsisArray}$ to empty
for $i \in [1, K]$ **do**
    $\boldsymbol{ConsisArray}[i] \Leftarrow$ CalConsist($\boldsymbol{S}, \boldsymbol{S^{Pi}}$)
end for
$P_{ci} \Leftarrow$ AVG($\boldsymbol{ConsisArray}$)
return P_{ci}

3.2. Motivation for Causal Intervention

For the link prediction task, the confidence score reflects the accuracy of the predicted triples. It is generally believed that high confidence has two meanings. The first is significance, that is, the obtained confidence score is optimal, which is significantly higher than other scores. The second is robustness, that is, the intervention embedding representation does not affect the measurement results.

Because the two confidence measurement methods mentioned in the related work only focus on the optimal score, in different sequences, the classification model has uncertainty in prediction, that is, it is difficult to accurately judge the credibility of the model output results. Therefore, this method starts from the aspect of robustness, and measures the performance of the model by conducting domain intervention on the embedded vector and comparing the robustness of the prediction results before and after the intervention. Inspired by the theory of causality [40], our approach is not to observe associations between different scores in a single sequence, but to judge the consistency of the output sequence by actively intervening in the scoring process by using multiple similar input vectors.

In causal inference, causal intervention refers to the method of confirming a causal relationship by changing a factor to observe its influence on other factors. Intervention can be active or passive changes, such as actively changing the value of a factor or natural events changing the value of a factor. In causal intervention, we hope to observe the influence of one or more independent variables on the dependent variables by changing one or more independent variables, and to verify whether there is a causal relationship between them. This process modifies a dimension in the input entity vector to exclude the influence of the variable on the scoring sequence.

To intervene in the scoring process, we first need to analyze the reasons that affect the generation of scoring

sequences. Given a trained KGE model M and multiple er queries with the same relationship r, the Structural Causal Model (SCM) [41] is used to describe the causal relationship in the KGE model. The entity vector e is an exogenous variable containing d dimensions, where each dimension value participates in the whole reasoning process as a variable. Through the calculation in M, the score sequence S is obtained. In this process, the transformed entity vector e' is the endogenous variable of e. Through the intervention method, it can help to judge the causality between the two, because the direct cause of the score sequence S. e' is determined not only by the variables of the same dimension in e, but also by the variables of other dimensions. We use causal intervention: $\mathrm{P}(S \mid e, \mathrm{do}(e_i))$. For d-dimensional embedding, we can generate d different neighborhood input vectors by modifying each dimension variable, the formal expression is described as follows.

$$\mathrm{P}(\boldsymbol{S} \mid \mathrm{do}(\boldsymbol{e})) = \frac{1}{d} \sum_{i=0}^{d} \mathrm{P}(\boldsymbol{S} \mid \boldsymbol{e}, \mathrm{do}(\boldsymbol{e}_i)) \tag{3}$$

3.3. Embedded Intervention

In the drug-target KG, triples are used to describe the specific DTI relationship, and they are stored and represented in the form of directed graphs. After model training, each entity and relationship will generate a unique embedded representation. This supervised learning process can better provide semantic features for embedded vectors and express the unique meaning of each entity. In the link prediction task, in order to more accurately determine the embedding vector representation corresponding to an entity, the intervention method is used for measurement and verification. For a drug-target entity, if it is represented by a d-dimensional embedding vector, then it has d features. We explore different intervention strategies to interfere with the feature values, resulting in d different intervention vectors by using the following strategy.

$$\boldsymbol{P_e} = (1 - \boldsymbol{I_d}) \times \boldsymbol{S_e} + \boldsymbol{I_d} \times \mathrm{ReplaceInterveVal}(\boldsymbol{e}) \tag{4}$$

where I_d is a d-dimensional unit matrix, and $ReplaceInterveVal(e)$ is to replace the dimensions in e with intervention values. The core idea of this method is to determine whether the original entity embedded in the representation has sufficient anti-interference and representativeness, and analyze it from the perspective of stability measurement.

Now, the method is more inclined to intervene in low-dimensional vectors, because high-dimensional data vectors are usually denser, which makes intervention and interpretation more challenging. If the number of dimensions is too large, it may lead to the transition of the vector from one semantic category to another semantic category, which will adversely affect the performance of the model and hinder the realization of the expected intervention results. Therefore, this paper only intervenes in the embedding representation value of one dimension in the experimental setting, and selects four relatively representative values, which are $\{0, \min, \max, \mathrm{AVG}\}$.

3.4. Domain Score Sequence Acquisition

Drug-target link prediction includes head entity prediction, relationship prediction, and tail entity prediction. This paper focuses on the prediction of the results of the target tail entity, that is, when a head entity and relationship are determined, the trained model is used to measure the confidence score of the tail entity. It is a multi-classification problem of screening among multiple targets. For the prediction task, it is necessary to replace each tail entity to form a new triplet, then the corresponding KGE model will get a set of scores containing different tail entities. Because the triplet with the highest score is not necessarily authentic, the confidence score is different from the traditional focus on the best score. We assess the confidence score by evaluating the robustness of the predicted intervention pseudo-entity.

To assess the consistency between the score sequence of the original entity and that of the domain entity, it is essential to document the entities associated with each score and their respective rankings. Therefore, it is more inclined to select the Top$K$ scores in the score sequence and their corresponding positions. The ranking index can be expressed as $\mathrm{ArgIndex}(S^{N_i}) = \mathrm{Argsort}(S^{N_i})$, where the Argsort() function is a commonly used array sorting function. It is mainly used to return the index value of the sorted array elements rather than the elements themselves.

As illustrated in Fig. 7, the drug entity is subjected to a causal intervention operation, which results in the creation of a set of domain intervention entities through modifications to the embedded representation values across various dimensions. The KGE model scores each triplet formed by the domain pseudo-entities and the corresponding relationships between the original entity and each target prediction entity. For example, a 4-dimensional embedded entity e, a relationship r, and n candidate target tail entities are known for $e - r$ query prediction. Due to the large number of tail entities, the tail entities with high scores of TopK are selected, and then the head entity embedding representation is

intervened. After the causal intervention, four domain pseudo-entities $E' = \left\{e'_1, e'_2, e'_3, e'_4\right\}$ are created. The model then scores the prediction triplet formed by each pseudo-entity e', relation r, and K tail entities T, generating the score sequence $S' = \left\{S'_1, S'_2, S'_3, S'_4 \cdots S'_K\right\}$. Next, the scoring sequence S'_i corresponding to each pseudo-entity e'_i and the scoring sequence S corresponding to the initial entity e are used to calculate the consistency difference, ultimately yielding a new consistency score.

3.5. Causal Intervention for Measuring Confidence

The consistency score, alternatively termed the confidence score, is assessed using the Confidence Interval (CI) in this study, signifying the reliability of entity embeddings. Illustrated in Fig. 8, the main method involves comparing and normalizing the consistency between the score sequences of pseudo-entities and the original entities across various fields. This process is predicated on the notion that for an entity deemed accurate within a triplet for link prediction, its embedding traits should correlate with the predicted scores by the model. Moreover, minor alterations to a single feature within an entity's embedding vector should minimally impact the entity overall. Thus, it becomes vital to evaluate the consistency across the scoring sequences.

For the domain sequence, the index ordering of the first K candidate entities is denoted as $\rho(S^{N_i})$, and the index robustness scores of the first K candidate entities in the initial score sequence are calculated as follows:

$$\boldsymbol{S_P} = \left\{\boldsymbol{S^{P_i}} | 0 < i \leq K, \boldsymbol{S^{P_i}} \in \mathrm{Sort}(\mathrm{Score}(\boldsymbol{e} \wedge \boldsymbol{P_e}))\right\} \quad (5)$$

$$\mathrm{Consis}(\boldsymbol{S}, \boldsymbol{S^{P_i}}) = \sum_{j=0}^{J} \mathrm{Softmax}(\boldsymbol{S})_j \cdot (1 - \mathrm{Sgn}(\left|\rho(\boldsymbol{S})_j - \rho(\boldsymbol{S^{P_i}})_j\right|)), \quad (6)$$

where J is a hyperparameter served as a threshold for determining the first J indexesfor fitting. A robustness score of 1 suggests that the position of the scores in the domain sequence aligns perfectly with those in the initial sequence. A score lower than 1 reflects greater discrepancies in the index rankings, especially within the top J positions, signifying a greater variance and reduced robustness. Following the manipulation of a feature, the robustness score is calculated using $\mathrm{Consis}(S, S^{P_i})$. Subsequently, for an entity embedding vector of dimension d, the confidence score is normalized in the following manner:

$$\mathrm{P_{ci}}(\boldsymbol{S}) = \frac{1}{d} \sum_{i=0}^{d} \mathrm{Consis}(\boldsymbol{S}, \boldsymbol{S^{P_i}}). \quad (7)$$

Since the range of the final $\mathrm{P_{ci}}(S)$ is in [0, 1], it is used as the confidence score of the KGE model. In this method, the judgment is based on the score ranking, and the relative size of the entity score sequence and the domain score sequence at each position is judged by comparing the Sgn function, thus avoiding the direct comparison of the score size, and using this implementation method can weaken the similarity between the entities. The possible impact of numerical differences in scores improves the stability and accuracy of the measurement.

3.6. Drug-Target Link Prediction Calibration

Model calibration is necessary for link prediction tasks. When the model uses the data set for embedded representation learning, the trained machine learning model is used to predict the corresponding DTI. However, the machine learning model often gives some uncertain prediction results when predicting. For example, some predicted probability values may be overestimated or underestimated. This will lead to a decrease in the reliability of the prediction results, thereby increasing R & D costs and misdiagnosis rates.

The process of calibration of the drug-target link prediction model is as follows. Firstly, in the model training stage, the embedded representation learning is carried out for the KG of the drug-target, and then the link prediction task is processed. When predicting the results, there will be a confidence score for each prediction result. To evaluate the reliability of the model more accurately, these confidence scores need to be calibrated to make the prediction results more accurate and reliable, and finally, the overall confidence scores need to be calibrated.

The specific operation step of calibration is that for a prediction result, its confidence score can be first normalized to the [0, 1] interval, and the interval is divided into n bins. Then, for each bin, the difference between the average confidence score of all the predicted results and their average true score is calculated. By weighting the scores of all bins, an ECE score can be obtained to evaluate the reliability of the model. The smaller the score is, the closer the confidence score of the model is to the real score, which means that the reliability of the model is higher. The use of ECE scores can help to more accurately evaluate the performance of the model and confirm whether the model has an over-confidence problem, thereby improving the reliability of the model.

$$\mathrm{ECE} = \sum_{b=1}^{B} \frac{n_b}{N} \left|\mathrm{acc}(\boldsymbol{b}) - \mathrm{conf}(\boldsymbol{b})\right|, \tag{8}$$

where b denotes the count of samples within a specific bin, acc(b) is the true average value of these samples, and conf(b) indicates the mean predicted value for the samples in this bin.

4. Experiments

4.1. Experimental Setup

In this experiment, three public knowledge maps, Hetionet, BioKG, and DRKG, were used in the field of drug-target discovery. In addition, 9 KGE models were trained based on the original settings. The dimension of the high-dimensional model is set to 200. For the low-dimensional model, the embedded dimension chosen in this paper is 50. For the hyperparameters of the CI, we use the entity of the former TopK with parameters 3, 5, 10, 100, 200, 300, and for the first J relative positions in the consistency calculation, the parameters are 3, 5, 10, 100. All models are implemented based on the Pykeen framework [42], and all algorithms are implemented using the PyTorch framework based on Python. All experiments were conducted on NVIDIA RTX A400 GPU and Intel(R) Xeon(R) Silver 4210R CPU @ 2.40GHz on CentOS.

4.2. Evaluation Measures

To evaluate the performance of the link prediction task, we introduce five evaluation indicators:

- **ECE**: Expected Calibration Error (ECE) is frequently used to assess model calibration; a lower ECE signifies a more effective calibration of the model.
- **ACC**: Triple accuracy (ACC) refers to the average accuracy of triples in the same confidence level.
- **T10ACC**: the average accuracy of the top 10% high-confidence triples. The higher the T10ACC, the better the confidence measurement effect of the confidence measurement method.
- **MRR**: Mean reciprocal ranking (MRR), the average value of the ranking of the entity pairs to be predicted in all possible entity pairs.
- **CVaR**: We rank the losses according to the probability distribution of the losses, and then take the average value of a portion of the losses at the tail at the confidence level. The estimation of the average loss in the worst case of the loss distribution.

4.3. Comparative Analysis of ECE and T10ACC

This study examined the impact of three confidence metrics on the accuracy of nine distinct KGE prediction models, comparing the Confidence Interval (CI) with the SIG and the TOP metrics. The Expected Calibration Error (ECE) and Top-10 Accuracy (T10ACC) scores were determined for each model across various datasets. For calibration in this experiment, two standard techniques were employed: Platt Scaling (Platt) [43] and Isotonic Regression (Isoto) [44].

Table 2 shows the ECE of different confidence measures for DTI prediction models. For ECE indicators, the SIG is generally higher than the TOP in the Hetionet, which is a multi-classification problem with tail entity prediction, which also shows that the Sigmoid function is more suitable for binary classification problems based on its characteristics, especially for low-dimensional models. It is more difficult to adjust the real confidence information flow, on the contrary, the TOP better reflects the advantages of multi-classification tasks, and the overall error of the model is smaller than that of the SIG. Compared with the other two methods, the CI has a more obvious effect, and obtains a lower ECE error value, indicating that the CI has achieved results in the stability of embedded representation, making the semantic representation of each entity and relationship more realistic. In the three data sets, the CI has a more obvious effect than the other two methods, and lower ECE error values are obtained in most of the KGE models, among which the three confidence levels in the DRKG data set are smaller than the ECE values in the other two data sets.

In Table 3 featuring Hetionet data, the CrossE model significantly enhanced low precision predictions, achieving a confidence accuracy of 0.521 as measured by CI, which suggests superior prediction accuracy over other models. Within both the Hetionet and DRKG datasets, the low ACC scores for most models contribute to poor overall confidence ratings. Specifically, the ACC for ComplEx on the Hetionet dataset surpasses that of BioKG. Although T10ACC scores derived from SIG and TOP methods are comparable, the CI values are nearly double, showing more pronounced benefits at higher ACC levels and less effectiveness at lower accuracies. Overall, in drug-target interaction prediction

Table 2 The experimental results w.r.t., ECE on different confidence measures on DTI predictions.

Model	Hetionet			BioKG			DRKG		
	ECE ↓			ECE ↓			ECE ↓		
	SIG	TOP	CI	SIG	TOP	CI	SIG	TOP	CI
TransE	.058	.032	**.016**	.150	.106	**.063**	**.005**	.023	.019
ComplEx	.094	.042	**.035**	.154	.088	**.027**	**.001**	.008	.006
DistMult	.051	.037	**.024**	**.026**	.062	.041	.004	.005	**.003**
RotatE	.035	.022	**.020**	.083	.093	**.063**	.006	.005	**.001**
TransH	.081	.027	**.013**	.034	.035	**.027**	.004	.006	**.003**
CrossE	.067	.037	**.021**	.018	.031	**.016**	**.013**	.014	.023
TransR	.051	.085	**.013**	.002	.001	**.001**	.003	.002	**.001**
HolE	.053	**.028**	.028	.008	.014	**.001**	.002	.002	**.002**
TuckER	**.001**	.034	.003	.004	**.002**	.005	.004	.005	**.001**

The optimal (minimum) results are in boldface.

Table 3 The experimental results w.r.t. T10ACC on different confidence measurement methods for DTI.

Model	Hetionet				BioKG				DRKG			
	T10ACC ↑			ACC ↑	T10ACC ↑			ACC ↑	T10ACC ↑			ACC ↑
	SIG	TOP	CI	H10	SIG	TOP	CI	H10	SIG	TOP	CI	H10
TransE	.146	**.237**	.121	.117	.453	.398	**.692**	.141	**.256**	.255	.255	.159
ComplEx	.022	.027	**.132**	.061	.015	.036	**.064**	.013	.002	.001	**.006**	.007
DistMult	.419	.301	**.428**	.168	.195	**.322**	.232	.053	.008	.001	**.039**	.013
RotatE	.299	.453	**.475**	.243	.397	.621	**.650**	.238	.005	.002	**.012**	.017
TransH	.278	.139	**.380**	.192	.455	.407	**.588**	.183	**.223**	.212	.222	.295
CrossE	.442	.249	**.521**	.154	.158	.227	**.394**	.066	.070	.111	**.147**	.145
TransR	**.254**	.025	.081	.081	**.021**	.003	.004	.003	**.003**	.002	.002	.003
HolE	.001	**.213**	.129	.121	.040	.020	**.068**	.057	.045	.042	**.053**	.055
TuckER	**.073**	.001	.045	.022	.014	.008	**.018**	.053	.004	.007	**.019**	.020

The optimal (minimum) results are in boldface.

tasks, a high T10ACC indicates that a model can more accurately predict correct drug-target pairs that align well with the model's high-confidence predictions, with a comparative analysis of T10ACC scores across models with varying embedding dimensions.

Therefore, the performance of the model with a low embedding dimension is more accurate. The embedded representation of relatively low dimensions can be preferred for experimental verification, which is also very important for the field of drug development, because the predicted results with high confidence are more likely to guide the design and development of subsequent experiments, and improve the efficiency of drug development.

4.4. Comparative Analysis of Confidence Measures

This experiment focuses on the RotatE model, aiming to delve into the impact of three different confidence measures on model performance under three calibration modes: Isotonic Regression, Uncalibrated, and Platt Scaling. Special attention is paid to the variations in ACC and ECE metrics corresponding to each confidence interval. To present the experimental results more intuitively, we employ reliability diagrams as a visualization tool, which uniformly divide the confidence range into 10 bins, each representing a specific confidence level interval. For each bin, we calculate the average predicted probability and actual probability of all samples within it. The horizontal axis reflects the average predicted probability, while the vertical axis represents the actual accuracy. By comparing these two values, we can assess the calibration degree of the model at different confidence levels. Specifically, within a confidence bin, the closer the average predicted probability is to the actual probability, the better calibrated the model can be considered at that confidence level. Conversely, a significant discrepancy between the two indicates that the model's calibration performance at that confidence level is suboptimal.

As shown in Figs. 2 a) - b), without any calibration method, the three confidence measurement methods are compared. Figure (a) is the Hetionet data set,

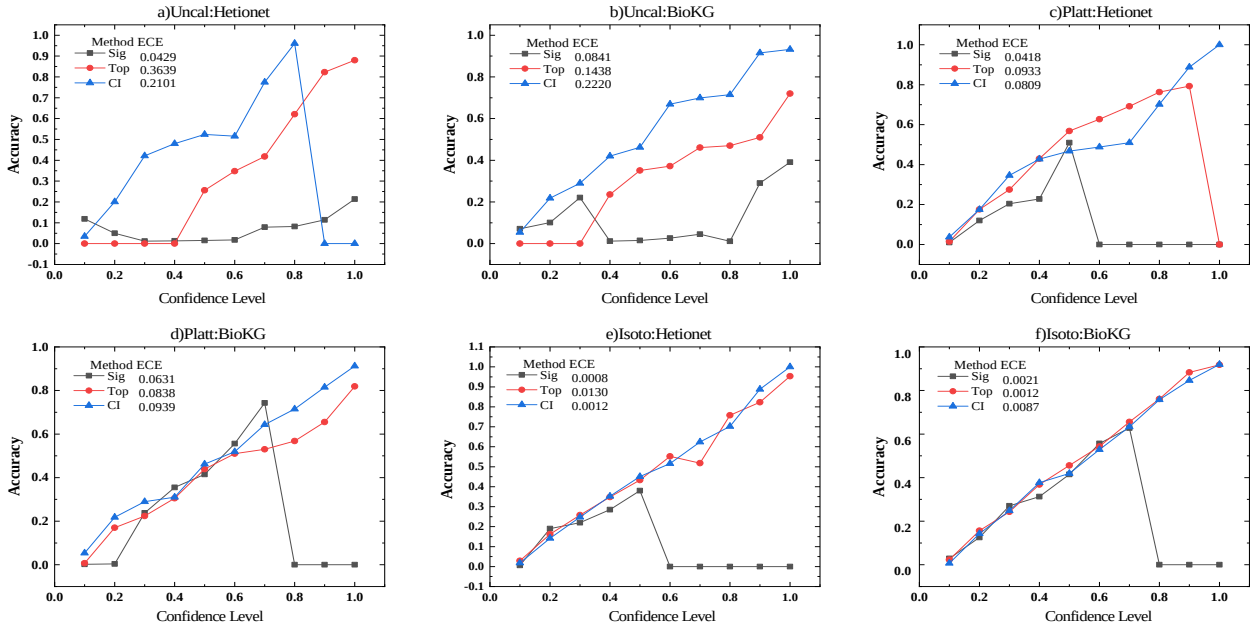

Fig. 2. Reliability diagrams for the RotatE model, using various confidence measurement and calibration methods across two datasets: Hetionet (a) and BioKG (b).

and Figure (b) is the Biokg data set. In the absence of calibration of the model, it can be seen that the SIG is more obvious in the high and low confidence scores, and the processing of multivariate classification tasks is limited. For extreme scores, the probability value may be close to 0 or 1, which may lead to excessive or insufficient confidence in the model. The TOP mainly focuses on the high confidence level, and there is no too low score. Compared with the SIG, it is more uniform, reflecting its superiority in multi-classification task processing. The CI can better predict the interaction relationship between drug-targets.

As shown in Figs. 2 c) - d), based on the Platt, it can be seen from the calibrated model that the ECE value has a significant decline, among which the SIG and the TOP have a significant decline, and the CI has a relatively small decline. Compared with Fig. 2 a) - b), it can be seen that the uncalibrated model is not true enough to predict the interaction between drug-targets, and the error is large. Especially for the traditional confidence measurement method, in the high confidence score interval, there are often untrue cases. Therefore, if the model is not calibrated while doing the drug discovery task, it can lead to a series of false negative and false positive predictions. Different from the SIG, the calibrated TOP returns the probability distribution in the low confidence interval to the normal range. In the Hetionet data set, the SIG is turned to zero after 0.5 after calibration, and in the BioKG data set, the threshold of convergence to 0 is larger. Therefore, it can be seen that in the Hetionet, the interaction prediction effect of the drug-target is slightly worse. Instead of using the CI, the confidence score is relatively uniform in the overall probability interval distribution, and the ECE decline is also relatively low before and after calibration, which also reflects the robustness of the CI compared with the traditional method.

The Isoto used to calibrate the model is shown in Figs. 2 e) - f). The model is calibrated based on the Isoto. It can be seen from the diagram that the overall calibration fit of the Isoto is better, especially in the TOP and CI, and the ECE index is also lower. Among them, the CI is also smoother than the TOP, which is close to the optimal calibration baseline.

Overall, the CI using the Isoto demonstrates superior performance on both Hetionet and BioKG datasets. Therefore, our proposed new method is validated from multiple perspectives to be superior to traditional confidence methods, particularly in the context of drug-target link prediction, where the results obtained after applying the Isoto are more authentic and reliable.

4.5. Single Hyperparameter Experimental Analysis

Subsection 3.5 mentioned that in the CI, when replacing each target entity in the prediction triple, there will be a corresponding score. Due to the large number of

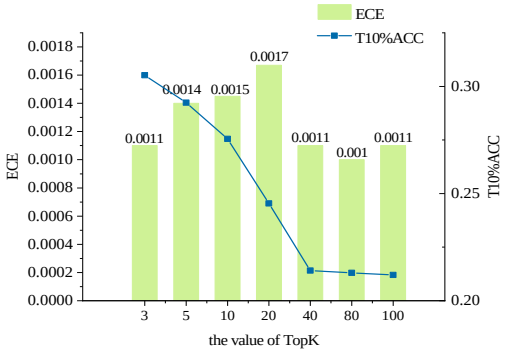

Fig. 3. Effect of hyperparameter K setting on confidence measurement

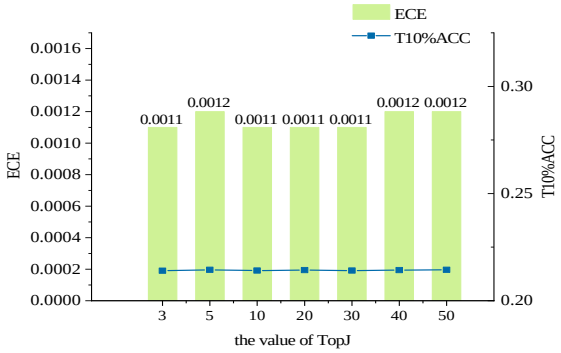

Fig. 4. Effect of hyperparameter J confidence measurement

tail entities, the tail entities with high scores in the former Top$K$ are selected, and then the head entity embedding representation is intervened. The corresponding Top$K$ tail entities in this part are selected for triple re-rating, and the first J intervention entities are selected as the measure of the number of consistent entities. Therefore, this experiment analyzes the impact of causal intervention results from changes in parameters. In this experiment, the DistMult and TransH models were selected to test on the Hetionet dataset and the BioKG dataset, respectively, to explore the effects of hyperparameters K and J on the CI, and their combined effects on model confidence scores. The indicators involved in the experiment include ECE, T10ACC, and CVaR. As indicated in subsection 3.5, the value of K is always greater than or equal to J, meaning the starting value of K varies in response to changes in J. In this experiment, the range for J and K is defined as $J \leq K \leq 300$.

The experiment first measures the influence of the hyper-parameter K on the confidence measurement results. From the above, it can be seen that the hyperparameter K represents the Top$K$ highest scores and their corresponding tail entities selected from the original entity score sequence. It represents how many pairs of drugs, targets, and the interaction between them are used in the domain entity prediction. The triples are scored and sorted. The following Fig. 3 indicates that when the hyperparameter J is a fixed value, the change of K is a double Y-axis columnar line combination diagram of the change distribution of ECE and T10ACC in the confidence measurement of the CI, where $J = 3$. The bar chart represents the value of ECE, and the line chart represents the value of T10ACC.

As shown in Fig. 3, this experiment explores the impact of hyperparameter K on ECE and T10ACC in measuring the CI, with hyperparameter J set to a constant value of 3 and K continuously increasing. A combined bar-line chart with dual Y-axes is plotted to visualize the distribution of ECE and T10ACC changes. Specifically, the bar chart represents the ECE values, while the line chart represents the T10ACC values. When K is less than 30, as K increases, T10ACC decreases and ECE increases overall. However, as K continues to increase, ECE remains stable at around 0.001, and T10ACC maintains a relatively low level. This indicates that when K is set to a smaller value, the performance of both ECE and T10ACC is better.

As shown in Figure. 4, this experiment investigates the impact of hyperparameter J on ECE and T10ACC in measuring the confidence of the CI, with hyperparameter K set to a constant value of 50 and J continuously increasing. The experimental results show that as J increases, there is almost no significant fluctuation in ECE and T10ACC. This indicates that during the hyperparameter selection process, the influence of parameter J on the model's confidence is relatively weaker compared to parameter K. Therefore, when adjusting model parameters, we should be more cautious in adjusting the value of K to ensure that the model can accurately capture the causal relationships between variables. For parameter J, on the other hand, an appropriate value can be selected based on the actual situation of computational resources without over-optimization. Such a strategy can balance the model's performance and computational cost, thereby improving its performance in practical applications.

4.6. Hyperparameter Threshold Analysis

It can be seen from subsection 4.5 that the parameter K plays a leading role in the CI. Therefore, the following two experiments will take the values of different J to explore the threshold selection of hyperparameter K. As shown in the following Figs. 5 and Fig. 6, it is the

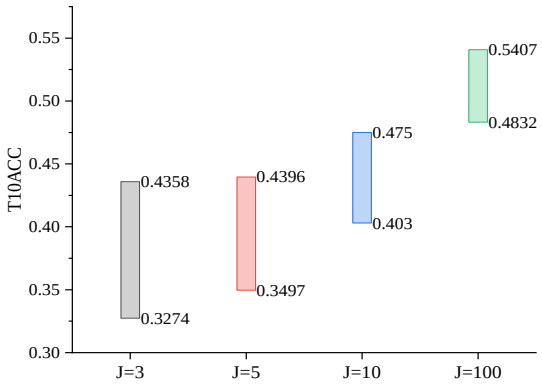

Fig. 5. Effect of hyperparameters K and J on T10ACC

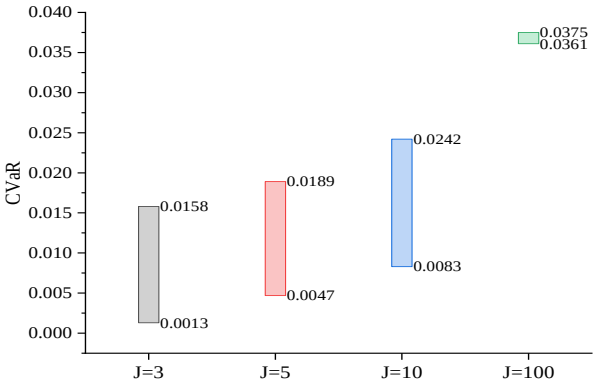

Fig. 6. Effect of hyperparameters K and J on CVaR

interval diagram of the influence of hyperparameters K and J on the confidence measurement index.

Among them, to eliminate the interference of the embedded dimension of the KGE model, this experiment selects the DistMult model to compare and verify on the Hetionet dataset. As shown in Fig. 5, it is the distribution diagram of the hyperparameters K and J on the T10ACC index interval. Among them, the lowest T10ACC value is shown at the bottom of each columnar interval. At this time, the hyperparameter $K = J$, the value above each columnar interval represents the highest T10ACC value, but it does not increase with the increase of K. Due to the limitation of computing power, when K is set to 300, there is a significant decrease, which also shows that T10ACC can only improve the accuracy of the phase high confidence triple.

By comparing the selection of different J, it can be found that T10ACC increases with the increase of parameter J and $K = J$, which proves that the consistency is more accurate when the selected intervention entity score sequence and the original score sequence have more values. However, with the continuous increase of K and J, it can be seen that the distribution range of T10ACC scores is shrinking, indicating that the improvement of hyperparameters has enhanced the ability of model prediction accuracy, but the influence is decreasing to a certain extent. Therefore, the appropriate choice of hyperparameters is the key to the effectiveness of the CI.

As shown in Fig. 6, it is the distribution map of the hyperparameters K and J on the CVaR index interval. CVaR represents the variance of the confidence score of the triple. In statistics, it is used to measure the degree of dispersion between the data in a data set and the average value of the data set. From the representation meaning of CVaR, it is used to evaluate the prediction quality of drug-target prediction models. If the variance

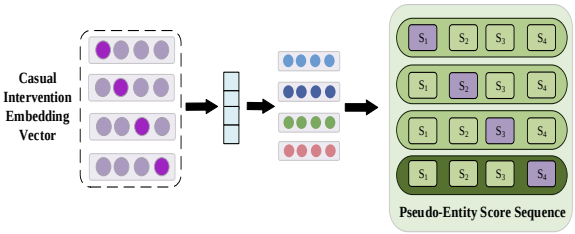

Fig. 7. Schematic diagram of pseudo entity score sequence

is relatively large, it shows that the difference between the prediction result and the training set is large, and the model needs to be further trained or optimized to improve its accuracy. Therefore, variance is an important indicator in drug-target prediction, which can help researchers evaluate the reliability and accuracy of the prediction results, to guide the drug development process. It can be seen from the figure that as J continues to increase, the value of CVaR is also increasing, indicating that the confidence of the triple is becoming more and more unstable. Especially when J is set to 100, the overall confidence score variance increases more obviously. Therefore, in the selection of hyperparameters, combined with the above tests, the selection of J cannot be too large, otherwise the accuracy measurement of the triple confidence score will be lost.

In the process of CVaR measurement, the highest value of $K = 300$ was also selected. With the increase of K, the change in CVaR value was small, because, in the process of measurement, the first J scores of the intervention entity score sequence and the original entity score sequence were selected. It plays a leading role in the process of index calculation and has a greater impact on the overall confidence variance.

In short, when considering the fractional stability of causal confidence, it is recommended that the value of K

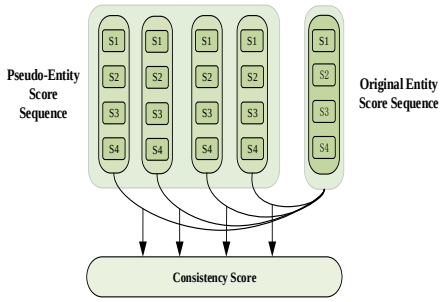

Fig. 8. Schematic diagram of consistency score calculation

is less than or equal to 100, and the value of J is less than or equal to 10. In the experiment, it is also found that when the average value of the high confidence triples is considered based on the overall model, the experimental analysis of the above two high and low dimension models is integrated. When the K value and the model embedding dimension value are approximately equal, the effect is the best. Therefore, this also greatly provides a powerful reference for the selection of consistent hyperparameters in the CI.

4.7. Causal Intervention Replacement Value Analysis

In addition to the selection of hyperparameters, this experiment also explored the effects of different intervention values on entity and relational embedding. Due to the wide range of intervention values to choose from, this paper sets it to four types: $\{0, \min, \max, \mathrm{AVG}\}$, so that the intervention is relatively representative, and the model selects two high-dimensional models DistMult and one low-dimensional model TransH that have not been calibrated, and still compares the two datasets to explore the experimental effects of different intervention values on ECE and MRR, where TransH(B) and DistMult(B) are represented on the Biokg dataset, and TransH(H) and DistMult(H) are represented on the Hetionet dataset.

Fig. 9 shows the impact of the CI on ECE, both the TransE model and the DistMult model, both of which have achieved relatively good results on the Hetionet dataset Under the setting of different intervention values when the intervention value is the maximum value of the embedding representation, the model calibration is relatively good from the perspective of different datasets and different models.

As shown in Fig. 10, the TransH model has a poor MRR measurement effect on the two datasets, similar to the change trend of ECE, when the average value and zero value based on the embedding vector are used for

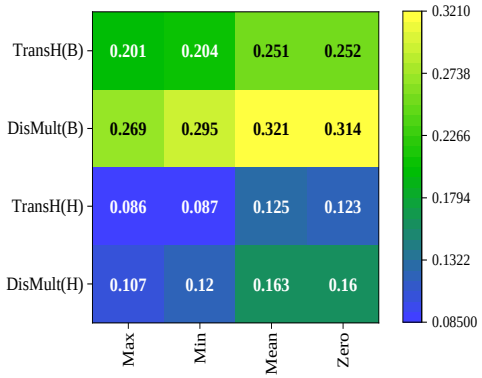

Fig. 9. Effect of causal interventions on ECE.

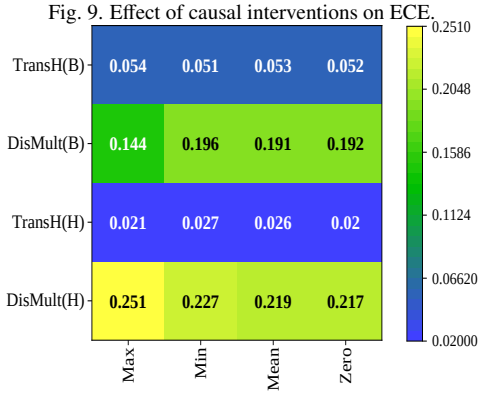

Fig. 10. Effect of causal interventions on MRR.

intervention, it will be found that the change range of ECE and MRR values is relatively stable, and when the maximum or minimum value is used for causal intervention value, the MRR is relatively high. Therefore, in the actual experiment, we should be more inclined to choose the maximum value to intervene in the dimension value.

5. Conclusion

This paper introduces a causal intervention confidence measure for DTI prediction. Specifically, three new KGE models (TransR, HolE, and TuckER) are designed to extend the comparison of the performance of different models and verify the robustness of previous experimental results, accompanied by a comprehensive introduction to causality supported by intricate formulas and pseudo-code. Additionally, the incorporation of a dataset broadens the experiment's scope, while extensive experiments analyze hyperparameter thresholds and scrutinize intervention substitution values.

The method performed well in the research, but the increase in computational cost during high-dimensional embedding poses a certain challenge for practical applications. In the future, it is necessary to explore di-

mensions that balance prediction accuracy and computational cost, as well as multi-dimensional intervention strategies and parameter optimization. Additionally, improving the consistency formula by incorporating score influence can enhance its universality and accuracy. Overall, this method serves as a beneficial supplement to enhancing model performance in the drug-target field, holding significant importance in expediting new drug research and development, reducing research and development costs, and improving drug safety.

Acknowledgment

This research project was supported by by National Key Research and Development Program of China under Grant 2023YFF1000100, and in part by the Major Project of Hubei Hongshan Laboratory under Grant 2022HSZD031, and in part by the open funds of Hubei Three Gorges Laboratory under Grant SK232011, and in part by Fundamental Research Funds for the Chinese Central Universities under Grants 2662023XXPY004, and in part by the open funds of the National Key Laboratory of Crop Genetic Improvement under Grant ZK202203, and in part by the Huazhong Agricultural University, and Inner Mongolia Key Scientific and Technological Project under Grant 2021SZD0099, and Australian Research Council under Grant DP230101122.

References

[1] A. Ezzat, M. Wu, et al., Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey, Briefings in Bioinformatics 20 (4) (2019) 1337–1357. 1

[2] N. Singh, P. Vayer, et al., Drug discovery and development: introduction to the general public and patient groups, Frontiers in Drug Discovery 3 (2023) 1201419. 1

[3] J. P. Hughes, S. Rees, et al., Principles of early drug discovery, British Journal of Pharmacology 162 (6) (2011) 1239–1249. 1

[4] R. Berwal, B. Bhyan, et al., A review on process of drug discovery and development, Annals of Biology (India) 28 (2). 1

[5] M. J. Keiser, V. Setola, et al., Predicting new molecular targets for known drugs, Nature 462 (7270) (2009) 175–181. 1

[6] E. Lounkine, M. J. Keiser, et al., Large-scale prediction and testing of drug activity on side-effect targets, Nature 486 (7403) (2012) 361–367. 1

[7] S. Núñez, J. Venhorst, C. G. Kruse, Target–drug interactions: first principles and their application to drug discovery, Drug Discovery Today 17 (1-2) (2012) 10–22. 1

[8] J. Chen, H. Peng, et al., Hogmmnc: a higher order graph matching with multiple network constraints model for gene–drug regulatory modules identification, Bioinformatics 35 (4) (2019) 602–610. 1

[9] X. Zeng, S. Zhu, et al., Network-based prediction of drug–target interactions using an arbitrary-order proximity embedded deep forest, Bioinformatics 36 (9) (2020) 2805–2812. 1

[10] L. Zhang, L. Zhang, et al., Mcdb: A comprehensive curated mitotic catastrophe database for retrieval, protein sequence alignment, and target prediction, Acta Pharmaceutica Sinica B 11 (10) (2021) 3092–3104. 1

[11] K. Ahmadi, F. Zahedifard, et al., Active site-based analysis of structural proteins for drug targets in different human coronaviruses, Chemical Biology & Drug Design 99 (4) (2022) 585–602. 1

[12] L. Hegazy, L. E. Gill, et al., Identification of novel mitochondrial pyruvate carrier inhibitors by homology modeling and pharmacophore-based virtual screening, Biomedicines 10 (2) (2022) 365. 1

[13] H. Öztürk, A. Özgür, E. Ozkirimli, Deepdta: deep drug–target binding affinity prediction, Bioinformatics 34 (17) (2018) 821–829. 1

[14] K. Wang, Y. Liu, Q. Z. Sheng, Neighborhood intervention consistency: Measuring confidence for knowledge graph link prediction., in: International Joint Conference on Artificial Intelligence, 2021, pp. 2090–2096. 1

[15] W. Ye, C. Li, et al., Causal intervention for measuring confidence in drug-target interaction prediction, in: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2023, pp. 395–400. 1

[16] A. Bordes, N. Usunier, et al., Translating embeddings for modeling multi-relational data, Advances in Neural Information Processing Systems 26. 2.1

[17] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, G. Bouchard, Complex embeddings for simple link prediction, in: International Conference on Machine Learning, PMLR, 2016, pp. 2071–2080. 2.1

[18] B. Yang, W.-t. Yih, X. He, J. Gao, L. Deng, Embedding entities and relations for learning and inference in knowledge bases, ArXiv Preprint ArXiv:1412.6575. 2.1

[19] Z. Sun, Z.-H. Deng, et al., Rotate: Knowledge graph embedding by relational rotation in complex space, ArXiv Preprint ArXiv:1902.10197. 2.1

[20] Z. Wang, J. Zhang, J. Feng, Z. Chen, Knowledge graph embedding by translating on hyperplanes, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 28, 2014. 2.1

[21] W. Zhang, B. Paudel, W. Zhang, A. Bernstein, H. Chen, Interaction embeddings for prediction and explanation in knowledge graphs, in: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, 2019, pp. 96–104. 2.1

[22] Y. Lin, Z. Liu, M. Sun, Y. Liu, X. Zhu, Learning entity and relation embeddings for knowledge graph completion, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 29, 2015. 2.1

[23] M. Nickel, L. Rosasco, T. Poggio, Holographic embeddings of knowledge graphs, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 30, 2016. 2.1

[24] I. Balažević, C. Allen, T. M. Hospedales, Tucker: Tensor factorization for knowledge graph completion, ArXiv Preprint ArXiv:1901.09590. 2.1

[25] L. Sleno, A. Emili, Proteomic methods for drug target discovery, Current Opinion in Chemical Biology 12 (1) (2008) 46–54. 2.2

[26] M. Abdullahi, S. E. Adeniji, et al., Homology modeling and molecular docking simulation of some novel imidazo [1, 2-a] pyridine-3-carboxamide (ipa) series as inhibitors of mycobacterium tuberculosis, Journal of Genetic Engineering and Biotechnology 19 (1) (2021) 12. 2.2

[27] S. Bonner, I. P. Barrett, et al., A review of biomedical datasets relating to drug discovery: a knowledge graph perspective, Briefings in Bioinformatics 23 (6) (2022) 404. 2.2

[28] H. Wang, G. Zhou, et al., Drug-target interaction prediction with graph attention networks, ArXiv Preprint ArXiv:2107.06099.

2.2
[29] F. Wang, X. Lei, et al., Predicting drug–drug interactions by graph convolutional network with multi-kernel, Briefings in Bioinformatics 23 (1) (2022) 511. 2.2
[30] Y. Yamanishi, M. Araki, et al., Prediction of drug–target interaction networks from the integration of chemical and genomic spaces, Bioinformatics 24 (13) (2008) 232–240. 2.2
[31] W. Zhang, Y. Chen, et al., Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data, BMC Bioinformatics 18 (2017) 1–12. 2.2
[32] Z. Yu, F. Huang, et al., Predicting drug–disease associations through layer attention graph convolutional network, Briefings in Bioinformatics 22 (4) (2021) 243. 2.2
[33] X. Li, Z. Xiong, et al., Deep learning for drug-drug interaction prediction: A comprehensive review, Quantitative Biology. 2.2
[34] L. Liu, F. Huang, et al., Multi-view contrastive learning hypergraph neural network for drug-microbe-disease association prediction, in: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, 2023, pp. 4829–4837. 2.2
[35] H. Y. Yuen, J. Jansson, Normalized l3-based link prediction in protein–protein interaction networks, BMC Bioinformatics 24 (1) (2023) 59. 2.2
[36] H. Wang, Z. Cui, Y. Yang, B. Wang, L. Zhu, W. Zhang, A network enhancement method to identify spurious drug-drug interactions, IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2.2
[37] R. K. Rungta, P. Jaiswal, B. Tripathy, A deep learning based approach to measure confidence for virtual interviews, in: International Conference on Computational Intelligence in Pattern Recognition, Springer, 2022, pp. 278–291. 2.3
[38] T. Pedro, C. Luca, Probability calibration for knowledge graph embedding models, ArXiv Preprint ArXiv:1912.10000. 2.3
[39] T. Safavi, D. Koutra, E. Meij, Evaluating the calibration of knowledge graph embeddings for trustworthy link prediction, ArXiv Preprint ArXiv:2004.01168. 2.3
[40] J. Pearl, D. Mackenzie, The book of why: The New Science of Cause and Effect, Basic Books, 2018. 3.2
[41] C. Hitchcock, Causality: models, reasoning and inference (2001). 3.2
[42] M. Ali, M. Berrendorf, et al., Pykeen 1.0: a python library for training and evaluating knowledge graph embeddings, Journal of Machine Learning Research 22 (82) (2021) 1–6. 4.1
[43] Platt, John, et al., Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, Advances in Large Margin Classifiers 10 (3) (1999) 61–74. 4.3
[44] G. Chuan, P. Geoff, et al., On calibration of modern neural networks, in: International Conference on Machine Learning, PMLR, 2017, pp. 1321–1330. 4.3