Transformed Graph Attention for Credit Rating

Charles Z. Liu[1], Sheng Xiang[1], Dawei Cheng[2,3*], Junyi Liu[2], Ying Zhang[1], Lu Qin[1]
[1]*AAII, University of Technology Sydney, Australia*
[2]*Department of Computer Science & Technology, Tongji University, Shanghai, China*
[3]*Key Laboratory of Artificial Intelligence, Ministry of Education, Shanghai, China*
[1]{charles.liu, sheng.xiang}@student.uts.edu.au;{ying.zhang,lu.qin}@uts.edu.au;
[2]{dcheng,1951805}@tongji.edu.cn

2023 IEEE 18th Conference on Industrial Electronics and Applications (ICIEA) | 979-8-3503-1220-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIEA58696.2023.10241546

***Abstract*—In banking finance and lending, credit rating is an important tool to measure creditor and debt reliability and prevent systemic risk among banks. Intrinsic features are key factors in risk control decisions. Decision models built with these intrinsic features are often used to assess risk classification and asset safety, but there lacks research on credit prediction for bank rating outlooks with financial correlation, which reflects the structured pattern of systemic risk and contagion. Although correlations among financial entities are valuable for rating outlook, it is hard to integrate such interactions into feature-based solutions. Aiming at the issues above, we propose a hyper-feature mapping operator based on an attention-based knowledge representation to build a knowledge system for the credit rating prediction of the interbank to represent the knowledge of credit rating. Combined with the interrelated behavior graph of each entity, a hyper context graph attention with intrinsic properties (assets, capabilities, buffers) are represented as knowledge for rating prediction to extract the context representation, helping understand the interaction of risk between intrinsic properties and entity-to-entity. Through the experimental comparison and evaluation based on 10 years of actual interbank data, the results show that the proposed model is effective with higher performance compared with the popular rating prediction methods currently used in the industry.**

***Index Terms*—interbank rating prediction, machine learning fintech, transformed graph attention representation; data mining**

I. Introduction

An interbank network is a network formed by inter-bank lending among commercial banks, in which the default of one bank may cause other banks to default, causing the risk of default to quickly spread to the entire network, and even lead to a systemic financial crisis [3], [6], [7]. Due to the obligatory relationships between banks, the credit rating between banks serves as a significant criterion to measure creditor and debt risk for financial reliability assessment [20]. Although the credit rating is intuitive and widely applicable, due to the fact that inter-bank financial shocks can evolve from small-scale contagion to systemic crises, and the rating has limited timeliness, it is still difficult for rating agencies and policymakers to detect and predict potential risks, thus formulating effective prudential supervision policies to prevent crises or conduct policy interventions. To handle this risk, rating outlooks [2], [8], [11] with a certain early warning mechanism are increasingly introduced into the credit rating process to assess the potential direction of a credit rating over a intermediate term thus preventing the latent risk before it occurs.

In the industry, the *domain rule based approaches* are designed by experts to predict the credit rating of the banks where various rules are manually designed based on the specific features of data and the mechanics of the interbank system. Despite of its importance in practice, as to our best knowledge, the problem of credit rating prediction has not been systematic studied for interbank system in the research literature. Nevertheless, similar to other finance related studies, the *classical machine learning models* can be directly applied without prior knowledge of rules or mechanism such as Regression Analysis (e.g., [12]), Decision Tree Model (e.g., [25]), Random Forest (e.g., [21]), gradient boosting decision tree (GBDT) (e.g., [31]), AdaBoost [18], Multilayer Perceptrons neural network (MLP) (e.g., [14], and Statistical Learning (e.g., [13], [16], [17], [23], [24]).

We observe that the domain rule based methods cannot work without proper hypothetical mechanism and may cause non-objective results in the rating due to the subjective assumptions of the model and the incomplete understanding of the mechanism. Moreover, these rules are usually not available for public or too specific due to the hand engineered nature of these methods. On the other hand, not surprisingly, our experiment results demonstrates that the performance of directly applying classical machine learning models is not competitive because they cannot well exploit the nature of the interbank data. Moreover, these models cannot properly capture the interconnection between entities (i.e., banks), while the lending relationship between interbank is complex, and the risk and rating of an entity will be affected by the status of other related entities.

In this paper, we model the interbank system as a network such that the interconnections among banks can be well captured. The prediction task on this network is quite challenging because of the complex nature of the data. The information about how much each bank owns in the interbank market usually could be available in the form of assets and liabilities, but the correlated influences among the banks are not easy to acquire and model directly (e.g., which bank borrowed money from which bank, the debt-creditor relationship and obligation extent between the two and other banks, etc.). Besides, feature

* Corresponding author: Dawei Cheng (dcheng@tongji.edu.cn)
This work was supported by the National Natural Science Foundation of China (Grant no. 62102287), the foundation of Key Laboratory of Artificial Intelligence, Ministry of Education, P.R. China and the Shanghai Science and Technology Innovation Action Plan Project (Grant no. 22YS1400600 and 22511100700)).

data, e.g., assets and liabilities, are heterogeneous, which means these features cannot be trivially fed into a neural network model.

To address the above challenges, in this paper we develop a transformer-based attention network, representing the correlations and influences with transformer-based graph attention representation (TGAR) as context knowledge. In particular, we use the feature data for the following three tasks: (1) generating the edges of the inter-bank relationship graph through a minimum density estimation, where the generated graph will be fed into our proposed method together with other input features; (2) deducing the strength of the inter-bank relationship by using a transformer-based network, where the self-attention mechanism can adaptively adjust the importance of each edge of the inter-bank graph; and (3) being passed on inter-bank relationship graph to infer the representation of each bank to predict their ratings. Moreover, as to our best knowledge, currently there is no public available interbank data which is a big hurdle for this line of research. In this project, we collect a inter-bank dataset spanning from 2010 to 2020, which will be public available for researchers. We conduct extensive experiments to evaluate the effectiveness of our proposed method on this dataset. The results demonstrate the superior performance of context attention representation-compared to baseline approaches.

We systematically address the problem of credit rating prediction for interbank system. Our work paves a new way of feature-driven predictive model in addressing the potential risk threat on inter-bank relationships. The corresponding design and implementation of the transformer-based graph attention representation enables the model to learn from the bank features directly. We also propose feature-driven self-attention mechanism in detail and prove its effectiveness in predicting the risk ratings. Comprehensive performance evaluation demonstrates the superior performance of our proposed methods compare to baselines. The results suggest that our method can help to prevent potential financial crisis in advance.

II. Method

A. Financial Background

The process of credit rating includes conference organization, Information Curation, Analysis & Assessment, Observation, sovereign national rating, deliberation, and rating determination [19] as shown in Figure 1. Different rating agencies have slightly different rating methods. The two largest rating agencies in the industry are Moody's and S&P Global Ratings [26]. At present, considering the changes in the market and economic conditions, rating outlooks [2], [8], [11] are increasingly introduced into the rating process as a supplement to the credit rating by the rating agencies' assessment of factors that may lead to changes in the rating review in the next 6-24 months, which is an essential application background of the topic in this paper.

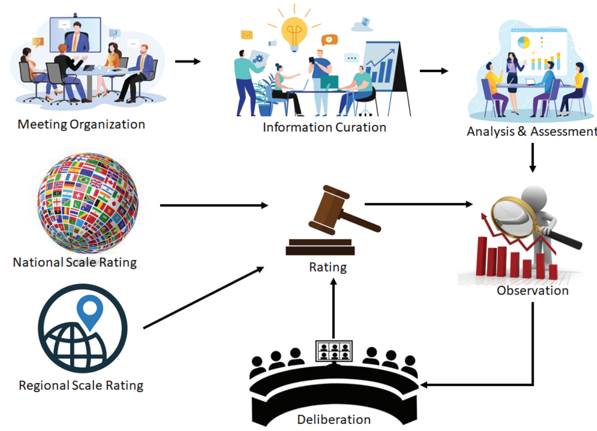

Fig. 1. Process of Credit Rating

B. Transformed Graph Attention Representation

Figure 2 shows an overview of the data used in this paper where the interbank data is represented by a graph. Particularly, each node (bank) is described by four attributes including *assets*, *capabilities*, *buffers*, and *weights*, and the entities are represented by X. Note that the weights of the entities are derived from the historical feature data of the banks by performing data fusion and the edges of the interbank graph are derived by a minimum density estimation [1], [3], [5], [22], which is represented by A. Note that the data architecture used in this paper does not require manual feature engineering and can be directly computed and learned by the proposed system.

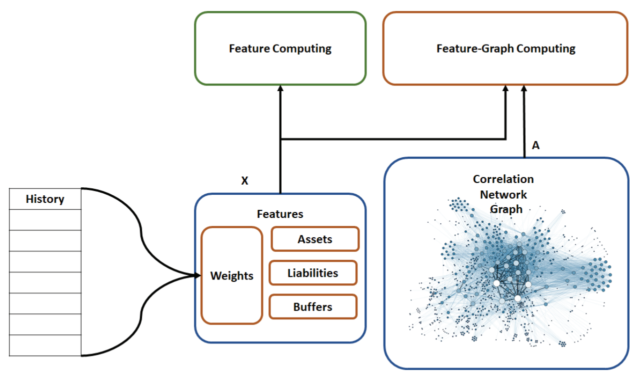

Fig. 2. Data Schema and Workflow

For the relationship between each entity, we use a graph to store the pointing relationship between each entity, and each pointing relationship represents the direction of the relationship behavior influence vector, which is encapsulated as A. The data architecture does not require manual feature engineering and constructs the original relational behavior into a sequential format of graph structure, which can be directly computed and learned by the proposed system.

We design a graph learning strategy with Transformed Graph Attention Representation (TGAR) to construct a knowledge representation of each entity for graph learning. Inspired by the work [29], [30], the Transformer-Based Graph Attention Representation (TGAR) can be described as shown in figure 3.

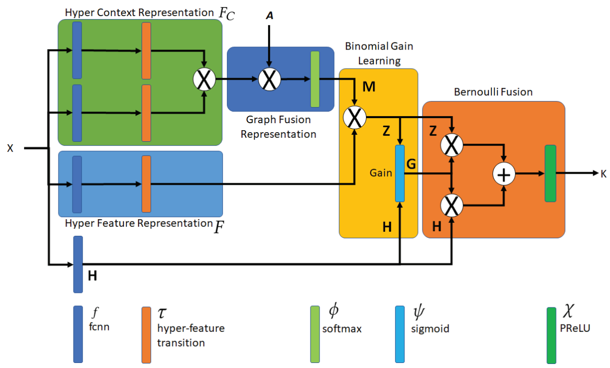

Fig. 3. Context Knowledge Representation with TGAR

C. Hyper Feature and Context Representation

We design a hyper feature unit and hyper context unit for extract the feature knowledge as a set of tensors. The main operators of hyper feature unit include $f(\cdot) = fcnn(\cdot)$ as a computation of a full connected neural network (FCNN) and $\tau(\cdot)$ as hyper-feature transition.

The FCNN operator is mainly used to establish a linear mapping with weights W that can be learned, and the mapping can be formulated as $f : X \longrightarrow y$ by

$$y = WX + b \tag{1}$$

where y refers to the mapping output and X input as a set of features, W and b the weight and bias to be learnt. In our design, we apply the FCNN modules to mapping the elementary features of X to functional projections combined with other operators, and these FCNN modules are the main parts of learning.

The hyper feature transition operator $\tau(\cdot)$ mainly performs the mapping from feature to tensor space, which can be formulated as

$$\tau : R_{n\times m} \longrightarrow R_{k\times n\times \frac{m}{k}} \tag{2}$$

in which k refers to the number of the hyper feature tensors. Therefore, the mapping of hyper feature unit $F(\cdot)$ can be formulated as

$$F(X) = \tau \circ fX \tag{3}$$

where $\circ$ represents operator multiplication, which is a cascade of inter-domain mappings (e.g. $H \circ F : A \longrightarrow B \longrightarrow C$ given $F : A \longrightarrow B, H : B \longrightarrow C$). The corresponding functional logic can be represented as $F(X) = \tau(f(X))$. As shown in figure 3 (b), the mapping of hyper context unit $F_C(\cdot)$ can be formulated as

$$F_C(X) = F_1(X) \times F_2(X) = \tau_1 \circ f_1 X \times \tau_2 \circ f_2 X \tag{4}$$

D. Graph Fusion Representation

As for the design of Graph Fusion Representation, a softmax [10] operator is used for multi-classification. Defining the softmax operator as $\phi(\cdot)$, and if the correlation can be represented as a graph A, the mapping of graph knowledge fusion can be formulated as

$$M(A, X) = \phi(AX) \tag{5}$$

where the softmax operator $\phi(\cdot)$ implements a mapping from a numerical space to a probability space. Its function in the system is to map the categorical numerical output to a relative probability as $P = \phi(C)$, by

$$p_i = \frac{e^{c_i}}{\sum_i^{N_c} e^{c^i}} \tag{6}$$

where $P = [p_1, p_2, \ldots]$ refers to the probability vector corresponding to the vector of classification output $C = [c_1, c_2, \ldots]$ in which c_i is the output of the ith classifier, the total number of categories is N_C, and p_i represents the probability ratio of the ith element as a probability over all possible classifications.

E. Binomial Gain Learning

We design a unit for a pair of tensor inputs to calculate the binomial gain, in which differential aggregation and sigmoid [9] operators are used to process the merging and mapping tensors as the gain, in which the sigmoid operator performs the mapping from the numerical space to the probability space, and can be formulated as $\psi(\cdot)$ where

$$\psi(x_i) = \frac{1}{1 + e^{-x_i}} \tag{7}$$

and the operation for the input in a vector form can be expressed as

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \Psi(X) = \begin{bmatrix} \psi(x_1) \\ \vdots \\ \psi(x_n) \end{bmatrix} \tag{8}$$

and the operation for the input in a matrix form $X = \{x_{ij}\}_{n\times m}$ can be expressed as

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \quad \Psi(X) = \begin{bmatrix} \psi(x_{11}) & \cdots & \psi(x_{1m}) \\ \vdots & \ddots & \vdots \\ \psi(x_{n1}) & \cdots & \psi(x_{nm}) \end{bmatrix} \tag{9}$$

The differential aggregation operator Υ mainly aggregate the inputs tensor as

$$\Upsilon(X_1, X_2) = f_{3\times n,1}([X_1, X_2, X_1 - X_2]^T) \tag{10}$$

where $f_{3\times n\times m,1}$ refers to a FCNN operator with an input size of $3 \times n$ and output size as 1.

Therefore, the binomial gain processing $G(\cdot)$ can be formulated as

$$G(X_1, X_2) = \Psi(\Upsilon(X_1, X_2)) \tag{11}$$

Although the softmax operator and sigmoid operator can map the tensor from a numeric domain to a probability domain, different from the use of the softmax in graph knowledge fusion, the purpose of using sigmoid in this differential aggregation is to distinguish the gains corresponding to the a pair of tensor inputs, while the softmax operator focuses on the probability mapping for multi-categorical outputs.

F. Bernoulli Fusion

As for the design of Bernoulli Fusion processing, a parametric rectified linear unit (PReLU) [28] operator is used for the activation of the output. Defining the PReLU operator as $\chi(\cdot)$, with the binomial gain G obtained by (11), the mapping of graph knowledge fusion can be formulated as

$$Y = \chi(GX_1 + (1 - G)X_2) \tag{12}$$

where $Y = [y_1, \ldots, _n]$ as the output. Let $\bar{X} = GX_1 + (1 - G)X_2 = [\bar{x}_1, \ldots, \bar{x}_n]$, it yields

$$y_i = \chi(\bar{x}_i) = \begin{cases} \bar{x}_i & \bar{x}_i > 0 \\ a\bar{x}_i & \bar{x}_i \leq 0 \end{cases} \tag{13}$$

in which $\bar{x}_i$ and y_i are corresponding to each other. In the paper, we set $a = 0.25$.

G. context attention representation mapping

Through the combination of the series of units and operators above, the context attention mapping can be completed under the systematic framework with hyper context unit and graph knowledge fusion on the input feature X and graph A. The graph context learning output M can be obtained by applying (4) and (5), which yields

$$M = \phi(A, F_C(X)) \tag{14}$$

With hyper feature unit, its output can be obtained by (3) as $F(X)$. After performing tensor multiplication with M, the product will be one of the inputs of binomial gain processing can be formulated as

$$Z = \phi(A, F_C(X))F(X) \tag{15}$$

Meanwhile, a hyper feature H is also obtained by $H = f_H(X)$ as another input to binomial gain processing. Together with Z, the binomial gain can be obtained by using (10),(11), which yields

$$G = \Psi(\Upsilon(H, Z)) \tag{16}$$

With H, Z, G the fusion representation K can be obtained by the Bernoulli fusion module, and the Bernoulli fusion procedure can be formulated as

$$K = \chi(GH + (1 - G)Z) \tag{17}$$

As the output of attention-based operations, the attention representation K can again be used as input for attention refinement, thus forming a cascaded context attention representation chain system.

TABLE I
ANNUAL STATS OF FEATURE DATA

Year	Avg Assets	Avg. Liabilities	Avg. Buffer	Avg. Weights
2011	10.1805	9.5426	326.8603	347.5834
2012	10.5614	9.8705	262.1356	283.5675
2013	10.7974	10.0724	498.7543	520.6241
2014	10.6820	9.9356	19795.6968	19817.3144
2015	10.4534	9.7032	27590.7611	27611.9178
2016	10.7571	9.9811	394.9608	416.6989
2017	11.8442	10.9607	1527.9176	1551.7226
2018	12.0645	11.1478	1981.3933	2005.6057
2019	12.2837	11.3360	1854.6793	1879.2990

III. EXPERIMENTS AND RESULTS

A. Data Set

In this section, we describe the extensive experiments for evaluating the effectiveness of different methods. The independent variables are 'assets', 'liabilities', 'buffer', 'weights' as input attributes, and the prediction target 'credit rank' of the next year as the output. From a variety of finance data sources such as Moody's Analytics BankFocus (https://bankfocus.bvdinfo.com/), Standard and Poor's Rating(https://www.spglobal.com/ratings/en/products-benefits/products/credit-ratings) and Trading Economics Credit Rating (https://tradingeconomics.com/country-list/rating), we collect 10-year interbank data from 2011 to 2020. The number of records in the attribute data is 14,245 each year; that is, there are 14,245 entities as prediction objects. The attributes contained in each entity are 'id', 'assets', 'liabilities', 'buffer', 'weights', 'credit rank', The stats of feature 'assets', 'liabilities', 'buffer', 'weights' is shown in table I. The attribute 'credit rank' include 4 rating categories, from the lower to the upper rank as D, C, B, A with the similar method as Moody's credit ranking [4].

B. Implementation and Evaluation

In this paper, we choose 2-layer cascading context attention representationto compose a cascading system and test its feasibility and performance. We apply Adam optimizer with $\beta_1 = 0.9, \beta_2 = 0.999$ and learning rate $\alpha = 0.02$. The experiment use the feature information of the previous year to predict the rating in the coming year, i.e., $y_{k+1} = h(X_t)$ in which X_t refers to the input of the year k, $h(\cdot)$ refers to the method used for prediction based on the input of year k, and y_{t+1} refers to the rating of the prediction. For methods with graph computing models, the model can be formulated as $y_{t+1} = h(X_t, A_t)$ where we use the feature input X_t and the graph input as A_t of year t to predict the rating of year $t + 1$.

To evaluate the performance of proposed method, we compare context attention representationwith the following models commonly applied in the risk assessment and prediction tasks, including the Ridge Classifier, Gradient Boosting Decision Tree (GBDT), Nearest Neighbors, Decision Tree, Random Forest, Neural Network, AdaBoost, Naive Bayes (NB), Multinomial NB, Bernoulli NB, Gaussian Process, and graph convolution network (GCN) [15], [27]. In order to evaluate the results of various methods from different perspectives, we choose four widely used metrics to measure the performance of different methods for credit rating prediction, including accuracy, average-weighted precision, macro-averaged recall, and macro-averaged F1-score.

Let $\{I_i\}$ be the aggregation of the indicators above. For different method $h_k(\cdot)$ and input in year t, the indicator can be formulated as $I_i(k,t)$, which refers to the ith indicator of the kth method in year t. To measure the expectation of metrics in the method-wise and yearly manner, we define the time-method-wise score $\bar{S}_k(t)$, metric-method-wise score $\bar{S}_k(i)$, and yearly score $\bar{S}_t(i)$ as

$$\bar{S}_k(t) = \frac{\sum_i I_i(k,t)}{M} \quad \bar{S}_k(i) = \frac{\sum_k I_i(k,t)}{T} \quad \bar{S}_t(i) = \frac{\sum_k I_i(k,t)}{K} \tag{18}$$

where M refers to the number of metrics, K the number of methods, T the number of years. In this paper, $K = 13, M = 4, T = 9$. These metrics reflect the performance expectation after synthesizing various t, k or i. To reflect the overall method-oriented performance, the comprehensive score of the kth method is introduced as

$$\bar{S}_k = \frac{\sum_i \sum_t I_i(k,t)}{MT} \tag{19}$$

This metric mainly focuses on evaluating the performance expectations of a method after synthesizing the metrics of the entire period. Similarly, the annual comprehensive score of the year t is defined as

$$\bar{S}_t = \frac{\sum_k \sum_i I_i(k,t)}{MK} \tag{20}$$

This metric mainly focuses on the synthesis of the performance metrics of each method in a certain period, showing the average level of prediction performance in this period with the implication of the expectation of difficulty in this period. The overall score is defined as

$$\bar{S} = \frac{\sum_i \sum_t \sum_k I_i(k,t)}{MTK} \tag{21}$$

This metric is a global performance metric expectation for the prediction task performance of the dataset itself after synthesizing time and method factors.

C. Main Results

According to the results of various experimental comparisons and evaluations, the context attention representationsystem that integrates graph-based context attention representation can perform more effective graph learning, and obtain more features under the same conditions to improve the prediction hit and coverage.

Figure 4 shows the comparison stats of metric comprehensive score. context attention representationmodel shows the highest $\bar{S}_k$ evaluation, followed by AdaBoost and Random Forest. Among the neural network methods, the performance of context attention representationis also much higher than that of GCN and MLP neural networks. Therefore, it can be seen that the attention representation integrated into the context has played a significant role in improving the performance of neural network methods. With (21), the overall metric $\bar{S}$ is 0.4349. Comparatively, the method-oriented performance of MLP neural network is lower than $\bar{S}$ due to the lack of correlation information, while GCN shows a boosting with the introduction of the graph, and context attention representation-improve more with the context attention. Figure 5 shows the

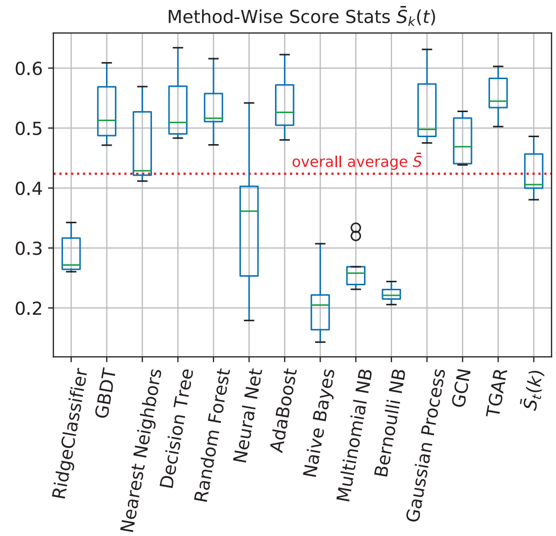

Fig. 4. Method-Wise $\bar{S}_k$ Comparison

yearly comparison between the synthesized metric $\bar{S}_k(t)$ and $\bar{S}_t$, revealing the difference between typical methods. In the forecast performance of each year, the performance expectation of Naive Bayes is always lower than the annual forecast average level, while the performance expectation of TGAR, GCN, GBDT is always higher than the average level, and context attention representationhas the highest performance value in the metric comparison of each year. The MLP neural network is below the annual average in most cases, only breaking the average for the two years in 2014 and 2017. This shows that when no graph information is introduced, the learning of MLP is more sensitive to the influence of the feature data of the current year, and the performance is not as good as the tree-based method. However, after the introduction of graph features, GCN and context attention representationare relatively stable in annual prediction performance, and the performance expectations can break through the annual average level.

IV. Conclusion and Discussion

In this paper, we propose a context attention representation predictive rating method for the application scenario of interbank rating outlook. We conduct experiments on our designed method on a real industry dataset and use 12 popular methods in the industry as baselines to compare the difference between methods. Experimental results show that proposed attention-based method outperforms these baselines in all metrics, and can have better performance in predicting the numerical features of expectation and variance. The advantages of our

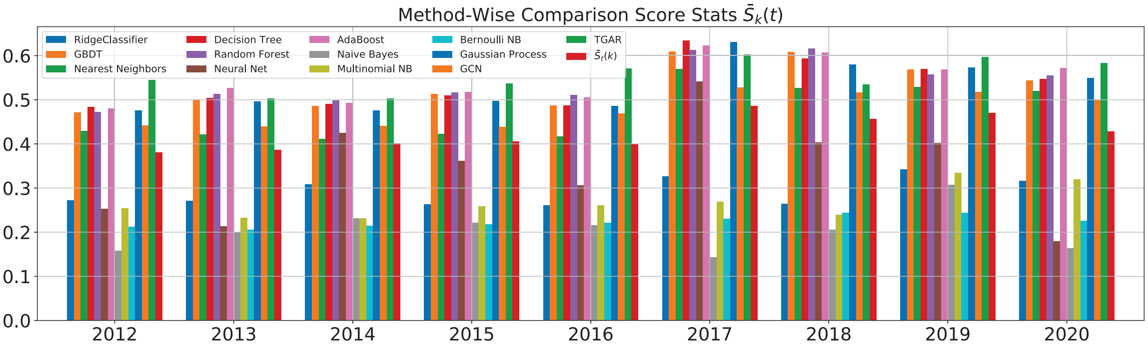

Fig. 5. Yearly Stats Comparison between $\bar{S}_t$ and $\bar{S}_k(t)$ of Typical Methods

solution are: 1) The weight embedding method under our data schema can significantly reduce the requirements on system storage space and computational resource consumption. 2) context attention representation maintains the independence of features, correlation, and feature-correlation while fusing heterogeneous feature data and relational data to reduce key information loss after information fusion. 3) The context attention mechanism can perform multiple learning on relation-feature, feature-feature with the Bernoulli fusion, and binomial gain. Future works include studying context gain and fusion mechanisms in other modalities, and their impact on prediction results in large-scale cascaded systems.

References

[1] Monitoring vulnerability and impact diffusion in financial networks. *Journal of economic dynamics control*, 76:109–135, 2017.

[2] Edward I Altman and Herbert A Rijken. The added value of rating outlooks and rating reviews to corporate bond ratings. In *financial management association meeting, Barcelona*, 2007.

[3] Kartik Anand, Ben Craig, and Goetz Von Peter. Filling in the blanks: Network structure and interbank contagion. *Quantitative Finance*, 15(4):625–636, 2015.

[4] J Richard Aronson and James R Marsden. Duplicating moody's municipal credit ratings. *Public Finance Quarterly*, 8(1):97–106, 1980.

[5] Marco Bardoscia, Stefano Battiston, Fabio Caccioli, and Guido Caldarelli. Debtrank: A microscopic foundation for shock propagation. *PLOS ONE*, 10(6), 2015.

[6] Michael Boss, Helmut Elsinger, Martin Summer, and Stefan Thurner 4. Network topology of the interbank market. *Quantitative finance*, 4(6):677–684, 2004.

[7] Giulio Cimini and Matteo Serri. Entangling credit and funding shocks in interbank markets. *PloS one*, 11(8):e0161642, 2016.

[8] John D Finnerty, Cameron D Miller, and Ren-Raw Chen. The impact of credit rating announcements on credit default swap spreads. *Journal of Banking & Finance*, 37(6):2011–2030, 2013.

[9] David John Finney. *Probit analysis: a statistical treatment of the sigmoid response curve*. Cambridge university press, Cambridge, 1952.

[10] Steven Gold, Anand Rangarajan, et al. Softmax to softassign: Neural network algorithms for combinatorial optimization. *Journal of Artificial Neural Networks*, 2(4):381–399, 1996.

[11] David T Hamilton and Richard Cantor. Rating transition and default rates conditioned on outlooks. *The Journal of Fixed Income*, 14(2):54–70, 2004.

[12] Barenya Bikash Hazarika, Deepak Gupta, and Parashjyoti Borah. An intuitionistic fuzzy kernel ridge regression classifier for binary classification. *Applied Soft Computing*, 112:107816, 2021.

[13] Shian-Chang Huang. Using gaussian process based kernel classifiers for credit rating forecasting. *Expert Systems with Applications*, 38(7):8607–8611, 2011.

[14] Xiaobing Huang, Xiaolian Liu, and Yuanqian Ren. Enterprise credit risk evaluation based on neural network algorithm. *Cognitive Systems Research*, 52:317–324, 2018.

[15] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[16] Huseyin Ozturk, Ersin Namli, and Halil Ibrahim Erdal. Modelling sovereign credit ratings: The accuracy of models in a heterogeneous sample. *Economic Modelling*, 54:469–478, 2016.

[17] Joonas Pajukoski. Predicting credit rating change using machine learning and natural lan. *Nature Methods*, 15(4):233–234.

[18] Kuldeep Randhawa, Chu Kiong Loo, Manjeevan Seera, Chee Peng Lim, and Asoke K Nandi. Credit card fraud detection using adaboost and majority voting. *IEEE access*, 6:14277–14284, 2018.

[19] S&P Global Ratings. S&p global ratings definitions. *URL: https://-www. standardandpoors. com/en_US/web/guest/article/-/view/sourceId/504352, Accessed on*, 15:13, 2016.

[20] Liliana Rojas-Suarez. Rating banks in emerging markets: What credit rating agencies should learn from financial indicators. In *Ratings, rating agencies and the global financial system*, pages 177–201. Springer, 2002.

[21] Miho Saito and Suguru Yamanaka. Performance evaluation of least-squares probabilistic classifier for corporate credit rating classification problem. *JSIAM Letters*, 13:9–12, 2021.

[22] Thiago Christiano Silva, Sergio Rubens Stancato de Souza, and Benjamin Miranda Tabak. Network structure analysis of the brazilian interbank market. *Emerging markets review*, 26:130–152, 2016.

[23] B Spoorthi, Shwetha S Kumar, Anisha P Rodrigues, Roshan Fernandes, and N Balaji. Comparative analysis of bank loan defaulter prediction using machine learning techniques. In *2021 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, pages 24–29. IEEE, 2021.

[24] Chih-Fong Tsai and Ming-Lun Chen. Credit rating by hybrid machine learning techniques. *Applied soft computing*, 10(2):374–380, 2010.

[25] Xianfu Wei. A method of enterprise financial risk analysis and early warning based on decision tree model. *Security and Communication Networks*, 2021, 2021.

[26] Lawrence J White. The credit-rating agencies and the subprime debacle. *Critical Review*, 21(2-3):389–399, 2009.

[27] Pingfan Xia, Zhiwei Ni, Hongwang Xiao, Xuhui Zhu, and Peng Peng. A novel spatiotemporal prediction approach based on graph convolution neural networks and long short-term memory for money laundering fraud. *Arabian Journal for Science and Engineering*, pages 1–17, 2021.

[28] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.

[29] Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform bad for graph representation? *arXiv preprint arXiv:2106.05234*, 2021.

[30] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. Graph transformer networks. *Advances in Neural Information Processing Systems*, 32:11983–11993, 2019.

[31] Yuhao Zhao. Research on personal credit evaluation of internet finance based on blockchain and decision tree algorithm. *EURASIP Journal on Wireless Communications and Networking*, 2020(1):1–12, 2020.