2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) | 978-1-6654-9402-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/WI-IAT55865.2022.00113

Modified Regularization for High-dimensional Data Decomposition

Sheng Chai
School of Computing
Queen's University
Kingston, Canada
s.chai@queensu.ca

Wenying Feng
Department of Computer Science
Trent University
Peterborough, Canada
wfeng@trentu.ca

Hossam Hassanein
School of Computing
Queen's University
Kingston, Canada
hossam.hassanein@queensu.ca

***Abstract*—With the increased dimensionality of datasets, high-dimensional data decomposition models have become essential data analysis tools. However, the decomposition method usually suffers from the overfitting problem and, consequently, cannot achieve state-of-the-art performance. This motivates the introduction of various regularization terms. The commonly applied Ridge regression has limited applicability for the asperity dataset and reduces performance for sparse data, while the Lasso regression has higher efficiency in the sparse dataset. To address this challenge, we propose a modified regularization term designed by integrating both the Lasso and Ridge regressions. The different roles of these two regressions are analyzed. By adjusting the weights of the regression in the regularization term, the existing decomposition method can be applied to the dataset with different degrees of sparsity. The experiments show that the modified regularization term yields consistent improvement in the performance of existing benchmarks.**

Keywords— high-dimensional data, regularization term, Lasso regression, Ridge regression, tensor decomposition

I. Introduction

With the increased dimensionality of real-world datasets, high-dimensional data decomposition model has become an essential data analysis tool. However, these decomposition models are negatively affected by the overfitting problem and, consequently, cannot achieve the state-of-the-art performance. This often requires regularization terms to enhance the decomposition performance. Commonly used regularization techniques are Lasso, Ridge, and Elastic Net regressions [1] [2]. Lasso regression uses the $L1$ norm, and the model that uses the $L2$ norm is called Ridge regression. Elastic Net regression, also called elastic network regression, combines Ridge regression and Lasso regression. Signoretto et al. extend the matrix norm to tensor data and applied it for supervised tensor learning to find low-rank projection matrices [3]. The success of the matrix trace norm inspireds Lacroix et al., and they propose a tensor p-norm regularization term [4] [5]. The Ridge regression is a popular regularization technique applied to the high-dimensional data decomposition modes [6]. It is mostly used to prevent overfitting when all features are extracted from the sample dataset [7]. However, experiments have shown that Ridge regression might reduce performance for sparse data, while the Lasso regression has higher efficiency in the sparse dataset [8]. It is desirable to consider a suitable regularization term to avoid overfitting issues.

In this paper, we propose a modified regularization term for high-dimensional data decomposition. The major novelty is the integration of two norms for estimating prediction results for sparse data recommendation, which simultaneously exploits Lasso and Ridge regression. The different roles of these two regressions are analyzed to solve the overfitting problem. The main contributions are:

1. Identify the Lasso and Ridge regression and find an effective way to help with high-dimensional data decomposition.

2. By combining the Lasso and Ridge regressions, the modified regularization term enhances the prediction performance and reduces the overfitting issue.

The rest of the paper is structured as the followings: Section I introduces the motivation and reviews some previous related work. Section II introduces tensor decomposition and regressions: Lasso, Ridge regression, and Elastic Net regression, respectively. Section III presents the modified regularization term to obtain the solution of the factor matrix in the tensor decomposition. Section IV applies the proposed algorithm to web service prediction with the experiment and analyzes the results. Finally, Section V summarizes the work.

II. Preliminaries

This section introduces the tensor decomposition, the notations of Lasso regression, Ridge regression, and Elastic Net regression method.

Tensor decomposition:

Tensor decomposition is a method for high-dimensional data analysis. For details of the algorithm, we refer can be found Kolda's related literature [9].

Given a 3-way tensor with rank-one $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_n}$ and an approximation tensor $\bar{\boldsymbol{\mathcal{X}}}$ corresponding to $\boldsymbol{\mathcal{X}}$, where $\mathbb{R}$ is the set of real numbers, I_n is the dimension of the tensor, and n is the number of dimensions. An approximation tensor $\bar{\boldsymbol{\mathcal{X}}}$ from decomposition satisfies the following equation (1),

$$\boldsymbol{\mathcal{X}} \approx \bar{\boldsymbol{\mathcal{X}}} = [\![U^{(1)}, U^{(2)}, \ldots, U^{(n)}]\!] \tag{1}$$

where $U^{(n)}\ n = 1,2,\ldots,n$ denotes the regular factor matrix. The symbol $[\![\bullet]\!]$ denotes the collection of factor matrices. The goal of the decomposition of $\boldsymbol{\mathcal{X}}$ is to find the regular factor matrix $U^{(n)}$ which produces the best approximation tensor $\bar{\boldsymbol{\mathcal{X}}}$.

Tensor decomposition is a complex model. Generally, a more complex model usually leads to overfitting. It might fail to predict future observations reliably [10]. To solve this problem, a regularization term needs to be added to the loss function to handle more complex learning tasks. The commonly used regularization terms are the Ridge regression of each factor matrix of a tensor.

Supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

978-1-6654-9402-1/22/$31.00 ©2022 IEEE
DOI 10.1109/WI-IAT55865.2022.00113

Authorized licensed use limited to: Queen's University. Downloaded on May 31,2023 at 18:30:34 UTC from IEEE Xplore. Restrictions apply.

Least absolute shrinkage and selection operator regression (Lasso):

Lasso regression is a regression model that uses the $L1$ norm $||W||_1$. It is defined as the sum of the absolute values of each element of the W. In feature selection, the $L1$ norm helps us minimize the objective function by making W equal to zero to remove these useless features and reduce the interference with the prediction of the sample. The objective function for the Lasso regression uses the $L1$ norm as shown in equation (2),

$$\|\mathcal{X}\|_1 = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3} |x_{i_1 i_2 i_3}|. \quad (2)$$

Ridge regression:

Ridge regression is a regression model that uses the $L2$ norm $||W||_2$. The $L2$ norm is a square root of the sum of the squares of the values W. The $L2$ norm makes each element of W small and close to zero. The smaller the parameter, the simpler the model, and the simpler it is, the less likely it is to produce overfitting. The Ridge regression solves the objective function, which is altered by adding a penalty equivalent to the square of the coefficients as shown in equation (3),

$$\|\mathcal{X}\|_2 = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3} x_{i_1 i_2 i_3}^2} = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}. \quad (3)$$

Elastic Net regression:

Elastic Net regression is a model that combines Ridge regression and Lasso regression.

III. Related Work

Commonly used regularization techniques include Lasso, Ridge, and Elastic Net regressions [11] [12]. Lasso regression uses the $L1$ norm, and the model that uses the $L2$ norm is called Ridge regression. Elastic Net regression, also called elastic network regression, combines Ridge and Lasso regression. Signoretto et al. extends the matrix norm to tensor data and uses it for supervised tensor learning to find low-rank projection matrices [13]. The success of the matrix trace norm inspires Lacroix et al., and they propose a tensor p-norm regularization term [14] [15]. The Ridge regression is a popular regularization technique applied to the tensor decomposition modes [16].

The Ridge regression is mainly used to prevent overfitting when all features are extracted from the sample dataset [17]. However, experiments show that Ridge regression might reduce performance for sparse data while the Lasso regression has higher efficiency in the sparse dataset [18]. Since the web service dataset is sparse, it is desirable to consider a suitable regularization term for the efficiency of the algorithms.

In recent studies, we also note that researchers use the N3 method to calculate the norm [19]. It is not suitable for a more general models. We have also conducted corresponding experiments, and the results show there are no significant difference between those using the N3 method to compute paradigms and ours.

IV. Methods

In this section, the solution for optimizing the loss function is introduced.

A. Basic regularization term

Computing an approximation tensor $\bar{\mathcal{X}}$ to satisfy $\mathcal{X} \approx \bar{\mathcal{X}}$ is essentially an optimization problem. The alternating least squares algorithm (ALS) is the most used optimization algorithm for tensor decomposition. The algorithm is applied in that one of the decomposition elements is optimized when other elements are kept fixed at each iteration. ALS estimates the factor matrices $U^{(n)}$ at each step by minimizing a loss function in the least squares.

For an original tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_n}$ and an approximation tensor $\bar{\mathcal{X}}$, the goal is to minimize the loss function L defined as shown in equation (4),

$$L = ||\mathcal{X} - \bar{\mathcal{X}}||^2 + \lambda \Omega(\bar{\mathcal{X}}) \quad (4)$$

where $\mathcal{X}$ denotes an original tensor, $\bar{\mathcal{X}}$ denotes an approximation tensor. $\lambda \Omega(\bar{\mathcal{X}})$ is the regularization term to avoid overfitting, which λ is the regularization parameter. $\Omega(\bar{\mathcal{X}})$ denotes regularization term. $|| \bullet ||$ denotes the norm of the tensor.

To avoid the overfitting issue, tensor decomposition commonly uses Ridge regression as the regularization term, which is shown in equation (5),

$$\Omega(\bar{\mathcal{X}}) = \frac{1}{2}\lambda \left(\left\|U^{(1)}\right\|^2 + \left\|U^{(2)}\right\|^2 + \cdots + \left\|U^{(n)}\right\|^2 \right) \quad (5)$$

where $\bar{\mathcal{X}}$ denotes an approximate tensor, $U^{(n)}\ n = 1,2, \ldots, n$ denotes the regular factor matrix. λ are parameters of the factor matrix in the regularization term.

B. Modified regularization term

The recommendation performance based on tensor decomposition is usually negatively affected by the overfitting problem and, consequently, cannot achieve state-of-the-art performance. This often requires regularization terms to enhance decomposition performance.

Motivated by the Elastic Net regression, we propose a modified regularization term that benefited the advantages of both Lasso and Ridge regressions as shown in equation (6):

$$\Omega(\bar{\mathcal{X}})_{new} = \lambda(\frac{1-p}{2} ||\bar{\mathcal{X}}||_2^2 + p||\bar{\mathcal{X}}||_1) \quad (6)$$

where $\bar{\mathcal{X}}$ denotes an approximation tensor, $\|\bullet\|$ denotes the tensor norm, $\lambda > 0$ is the regularization parameter. The parameter $p = 0$ corresponds to the ridge method $\|\bullet\|_2$ and $p = 1$ to the lasso method $\|\bullet\|_1$. $\boldsymbol{\lambda}$ denotes the regularization parameter, and its default value is 35. $\Omega(\bar{\mathcal{X}})_{new}$ denotes modified regularization term.

An optimization task for the objective function $L(\mathcal{X}, \bar{\mathcal{X}})$ is performed as shown in equation (7),

$$\min L(\mathcal{X}, \bar{\mathcal{X}}) = \min(||\mathcal{X} - \bar{\mathcal{X}}||^2 + \lambda \Omega(\bar{\mathcal{X}})_{new})$$

$$\Rightarrow \min(||\mathcal{X} - \bar{\mathcal{X}}||^2 + \lambda \left(\frac{1-p}{2} ||\bar{\mathcal{X}}||_2^2 + p||\bar{\mathcal{X}}||_1\right)$$

$$\Rightarrow \min(||\mathcal{X} - \bar{\mathcal{X}}||^2) + \lambda \left(\frac{1-p}{2}\left(\left\|U^{(1)}\right\|^2 + \left\|U^{(2)}\right\|^2 + \cdots + \left\|U^{(n)}\right\|^2\right) + p\left(\left\|U^{(1)}\right\|^2 + \left\|U^{(2)}\right\|^2 + \cdots + \left\|U^{(n)}\right\|^2\right)\right) \quad (7)$$

Authorized licensed use limited to: Queen's University. Downloaded on May 31,2023 at 18:30:34 UTC from IEEE Xplore. Restrictions apply.

where $U^{(n)}$ $n = 1,2, \ldots, n$ denotes the regular factor matrix.

The regularization can be treated as a compromise between finding a small penalty and minimizing the loss function $L(\mathcal{X}, \bar{\mathcal{X}})$. The regularization parameter λ controls the compromise: the smaller the λ., the more it minimizes the loss function, and conversely, the smaller the penalty.

The setting of the regularization parameters is related to the size of the dataset. Usually, the regularization parameters are set larger for large datasets and smaller for small datasets. To facilitate comparison with other methods, we use the default value $\lambda = 35$ in the experiment.

C. Algorithm

The tensor decomposition algorithm is given as the following Algorithm. Each iteration computing is performed in two steps: computing the feature factor matrices and updating the iteration.

Algorithm: Tensor decomposition with a modified regularization term

Input: an original tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_n}$, the regularization term parameter λ, the weight parameter p.

Output: an approximate tensor $\bar{\mathcal{X}}$.

Step 1. Initialize regular factor matrices $U^{(n)}$ $n = 1,2, \ldots, n$.

Step 2. Fixing the $U^{(2)}, U^{(3)}, \ldots, U^{(n)}$ to estimate the factor matrices $U^{(1)}$, $\overline{U^{(1)}}$.

Step 3. Compute the corresponding regularization $\lambda\Omega(\bar{\mathcal{X}})$, update the approximate tensor $\bar{\mathcal{X}}$.

Step 4. Reduce the objective function $L(\mathcal{X}, \bar{\mathcal{X}}) = \|\mathcal{X} - \bar{\mathcal{X}}\|^2 + \lambda\Omega(\bar{\mathcal{X}})$.

Step 5. Fixing the $U^{(1)}, U^{(2)}, \ldots, U^{(n-1)}$ to estimate the factor matrices $U^{(n)}$, $\overline{U^{(n)}}$.

Step 6. Compute the corresponding regularization $\lambda\Omega(\bar{\mathcal{X}})$, update the approximate tensor $\bar{\mathcal{X}}$.

Step 7. Reduce the value of objective function $L(\mathcal{X}, \bar{\mathcal{X}}) = \|\mathcal{X} - \bar{\mathcal{X}}\|^2 + \lambda\Omega(\bar{\mathcal{X}})$.

Step 8. Repeat the above steps and update the approximate tensor $\bar{\mathcal{X}}$ until convergence is exhausted.

Step 9. Return the final approximation tensor $\bar{\mathcal{X}}$.

The algorithm fixes the $U^{(2)}, U^{(3)}, \ldots, U^{(n)}$ to find the factor matrix $U^{(1)}$ and compute the approximate tensor $\bar{\mathcal{X}}$. If the convergence criterion $minL(\mathcal{X}, \bar{\mathcal{X}})$ is not satisfied, the algorithm continues to fix $U^{(1)}, U^{(3)}, \ldots, U^{(n)}$ to calculate the matrix $U^{(2)}$. The iteration continues to repeat the same procedure, which fixes $U^{(1)}, U^{(2)}, \ldots, U^{(n-1)}$ to find the matrix $U^{(n)}$. The convergence within the same loop requires multiple iterations until the convergence criterion is satisfied.

V. EXPERIMENTS

In this subsection, we implement the prediction experiments on the web service dataset to evaluate the novel regularization term for tensor decomposition.

A. Experimental settings

To evaluate the proposed QoS attribute prediction method, we use the web service dataset offered by Zheng et al. [20]. This dataset describes real-world QoS attribute prediction results from 142 users on 4,500 web services over 64 different time slices. This experiment focused on the response time and proposed a method to predict missing QoS attribute values.

The experiment is conducted on a Lenovo ThinkCentre M58 desktop with a 3.0 GHz Intel Core™ 2 Duo CPU and an 8 GB RAM, running an Ubuntu operation system. The program is implemented with Python 3.4 and Microsoft C++.

We use the standard mean absolute error (MAE), and root mean square error (RMSE) to compare the quality of our prediction.

MAE is defined as shown in equation (8):

$$MAE = \frac{\sum_{i,j} |r_{i,j} - \hat{r}_{i,j}|}{N} \quad (8)$$

RMSE is defined as shown in equation (9):

$$RMSE = \sqrt{\frac{\sum_{i,j} (r_{i,j} - \hat{r}_{i,j})^2}{N}} \quad (9)$$

where $r_{i,j}$ denotes the expected QoS attribute of web service j observed by user i, $\hat{r}_{i,j}$ is the predicted QoS attribute, and N is the number of the predicted value.

We verify the effectiveness of the proposed method, and the comparison is based on service collaboration with the following other methods.

Web service QoS attribute prediction framework (WSPred):

As a benchmark tensor decomposition method, this is a recommendation with a time-aware personalized QoS attribute prediction service for different service users [20]. WSPred uses Ridge regression as the regularization term and $\boldsymbol{\lambda} = 35$ as the default value.

Traversal-tensor method (TTM):

TTM remedies the shortage of low prediction accuracy rates caused by the lack of initial data samples. The method integrates the feature factor matrices to construct more data samples for tensor decomposition in our previous work [21]. When $\boldsymbol{\lambda} = 0$, this tensor decomposition method is executed without the regularization term.

TTM with regularization term method (TTMwR):

Based on TTM, our research employs the modified regularization term, which integrates the Lasso and Ridge regressions.

The above methods predict the response time and compute the MAE and RMSE values. The smaller value means the method has high performance.

Since a user does not revoke all web services, the dataset is usually sparse in the real world. The implement will randomly remove QoS attribute with different density from 5%, 10%,15%, 20%, 25%, and 30%. The 5% density means that 5% of the data is used for training, and 95% of the data is used for testing. We randomly set the parameter p in TTMwR corresponding to the Lasso and Ridge regressions. For example, the parameter $p = 0$ is to the Ridge, $p = 1$ to the lasso, and $p = 0.25$ means that result is generated by combing 25% by Ridge and 75% by Lasso regression.

Authorized licensed use limited to: Queen's University. Downloaded on May 31,2023 at 18:30:34 UTC from IEEE Xplore. Restrictions apply.

B. *Main results*

We set the different parameter $\boldsymbol{\lambda}$ value: when setting $\boldsymbol{\lambda}$ default value is 35, the TTM has the regularization term. When $\boldsymbol{\lambda} = 0$, it means that TTM has no regularization term. TTMwR shows better predictive performance than TTM without regularization and WSPred methods in TABLE I and TABLE II.

TABLE I. PERFORMANCE COMPARISON IN MAE

Methods	λ	P	Density 5%	Density 10%	Density 15%	Density 20%	Density 25%	Density 30%
WSPred	35		0.7913	0.7603	0.7535	0.7629	0.7520	0.7687
TTM	0		0.8183	0.7745	0.7417	0.7415	0.7345	0.7382
TTMwR	35	0	**0.6850**	0.6806	0.6723	0.6683	0.6604	0.6693
		0.25	0.6892	0.6721	0.6663	0.6672	**0.6594**	0.6763
		0.5	0.6859	0.6695	0.6721	**0.6665**	0.6635	0.6629
		0.75	0.6872	0.6711	0.6680	0.6682	0.6667	**0.6619**
		1	0.6884	**0.6679**	**0.6675**	0.6676	0.6647	0.6670

TABLE II. PERFORMANCE COMPARISON IN RMSE

Methods	λ	P	Density 5%	Density 10%	Density 15%	Density 20%	Density 25%	Density 30%
WSPred	35		1.8006	1.7741	1.7695	1.7764	1.7780	1.7823
TTM	0		1.8569	1.7852	1.7408	1.7366	1.7226	1.7214
TTMwR	35	0	1.5891	1.5788	1.5687	1.5680	1.5609	1.5645
		0.25	1.5989	**1.5710**	1.5669	1.5637	1.5617	1.5637
		0.5	1.5959	1.5721	1.5681	1.5635	1.5614	**1.5590**
		0.75	**1.5950**	1.5738	1.5673	1.5643	1.5633	1.5595
		1	1.6010	1.5725	**1.5666**	**1.5631**	**1.5601**	1.5621

C. *Accuracy comparison*

The TTMwR method has smaller MAE and RMSE values for all densities than the other methods in Fig. 1 and Fig. 2. The prediction accuracy can also be improved with the training matrix density increase from 5% to 30%. The total average MAE of the TTMwR method (0.67) has 14% more than the WSPred method (0.7648). Thus, the TTMwR method can significantly improve the accuracy result.

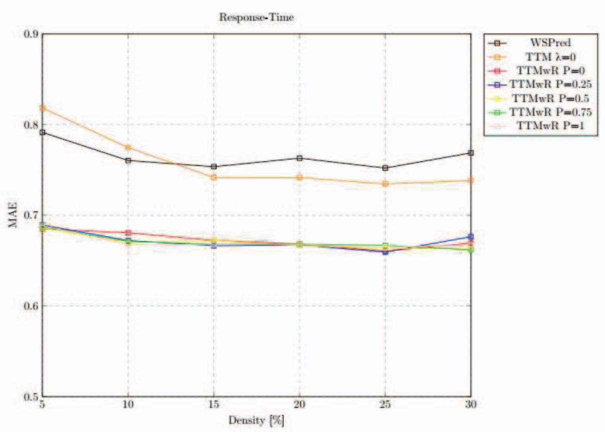


Fig. 1. Impact of density on prediction accuracy MAE

We also illustrate the evaluation results in the different values of the parameter p separately in Fig. 3 and Fig. 4 for the TTMwR method. The result shows that the method has the worst prediction accuracy for 5% density. With density increasing, the accuracy performance curve drop-down from 5% to 25% density. The best accuracy result appears when the density is 25% for both MAE and RMSE, and the MAE result generates a sharp decline curve at the point. Then the curve rises slightly after 25% density and the accuracy decrease for 30% density.

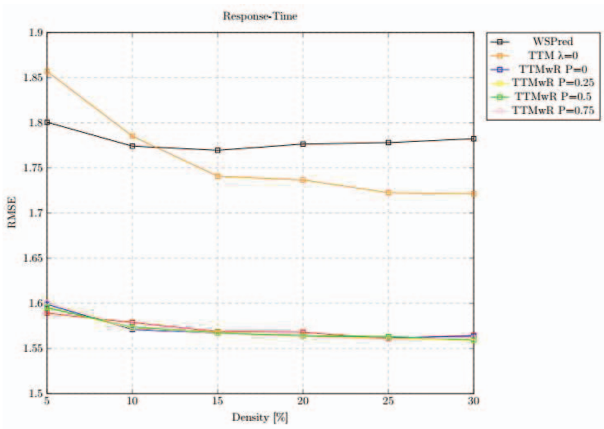
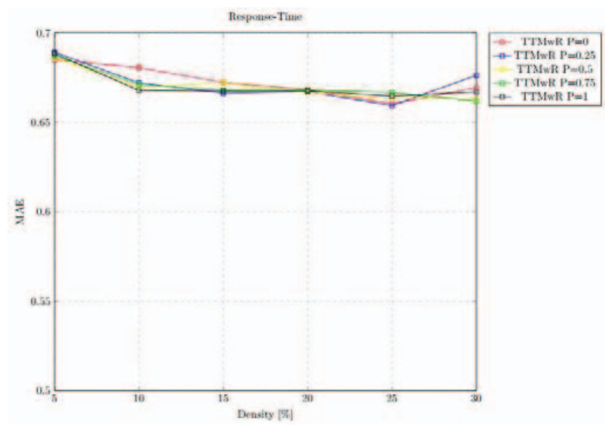


Fig. 2. Impact of density on prediction accuracy RMSE

D. *Weight impaction comparison*

We focus on the impact of the regression parameter p for the TTMwR method in Fig. 3 and Fig. 4.

The high MAE result is shown In Fig. 3 for a density 10% when $p = 1$, which means the Lasso regression help to achieve a better MAE result. As well as the best RMSE result appears when $p = 0.25$. Conversely, the minimum MAE result is shown when $p = 0.25$, and the minimum RMSE result is shown when $p = 1$ from a density of 15% to 25%. For density 30%, the best MAE result is shown when $p = 0.75$, and the best RMSE result is shown when $p =$



$0.5\ or\ 0.75$.

Fig. 3. mpact of the parameter p on MAE

Authorized licensed use limited to: Queen's University. Downloaded on May 31,2023 at 18:30:34 UTC from IEEE Xplore. Restrictions apply.

For the lower data densities (less than 25%), with the parameter p increases, the accuracy value rises. The Ridge regression contributes more to improving the performance of the method. However, when data density is higher, the performance depends on the contributions of the Lasso and Ridge regression. The best regularization ratio is 25% by the Lasso and 75% by the ridge, or an equal split.

Thus, from the experimental results, it can be observed that the impact of the parameter p is insignificant.

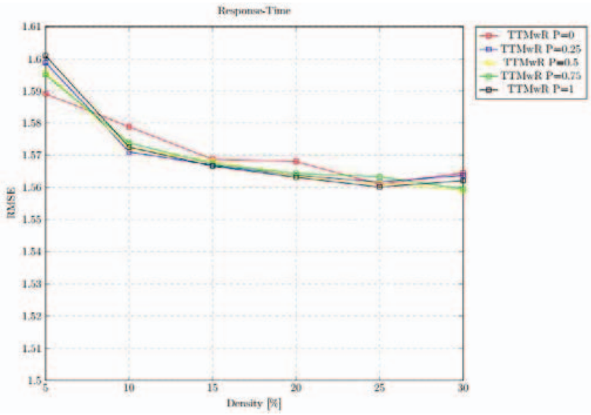


Fig. 4. Impact of the parameter p on RMSE

VI. Conclusion And Future Work

A modified regularization term incorporating two basic regressions for tensor decomposition is proposed. This method aims to reduce the possibility of overfitting and enhance the recommendation performance. Experimental validation is conducted on the web service dataset to discuss and evaluate the weights of the different regressions. The results show that the modified regularization term can effectively improve estimation performance and support web service prediction. In the future, conducting more evaluation on new data sets is essential, especially under various network environment scenarios [22] [23].

Acknowledgment

The work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). We are grateful to anonymous reviewers who have contributed to the enhancement of the paper's completeness with their valuable comments

References

[1] G Ogutu, Joseph O., Torben Schulz-Streeck, and Hans-Peter Piepho. "Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net, and their extensions." In BMC proceedings, vol. 6, no. 2, pp. 1-6. BioMed Central, 2012.

[2] Ji, Yuwang, Qiang Wang, Xuan Li, and Jie Liu. "A survey on tensor techniques and applications in machine learning." IEEE Access 7 (2019): 162950-162990.

[3] Signoretto, Marco, Lieven De Lathauwer, and Johan AK Suykens. "Nuclear norms for tensors and their use for convex multilinear estimation." Submitted to Linear Algebra and Its Applications 43 (2010).

[4] Lacroix, Timothée, Nicolas Usunier, and Guillaume Obozinski. "Canonical tensor decomposition for knowledge base completion." In International Conference on Machine Learning, pp. 2863-2872. PMLR, 2018.

[5] Candès, Emmanuel J., and Benjamin Recht. "Exact matrix completion via convex optimization." Foundations of Computational mathematics 9, no. 6 (2009): 717-772.

[6] Nickel, Maximilian, Volker Tresp, and Hans-Peter Kriegel. "A three-way model for collective learning on multi-relational data." In Icml. 2011.

[7] Zhang, Jianguang, Yahong Han, and Jianmin Jiang. "Tucker decomposition-based tensor learning for human action recognition." Multimedia Systems 22, no. 3 (2016): 343-353.

[8] Ruffinelli, Daniel, Samuel Broscheit, and Rainer Gemulla. "You can teach an old dog new tricks! on training knowledge graph embeddings." (2020): 1-12

[9] Kolda T.G. "Multilinear operators for higher-order decompositions." Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA (United States); 2006 Apr 1.

[10] Hawkins, Douglas M. "The problem of overfitting." Journal of chemical information and computer sciences 44.1 (2004): 1-12.

[11] Ogutu, Joseph O., Torben Schulz-Streeck, and Hans-Peter Piepho. "Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions." BMC proceedings. Vol. 6. No. 2. BioMed Central, 2012.

[12] Ji, Yuwang, et al. "A survey on tensor techniques and applications in machine learning." IEEE Access 7 (2019): 162950-162990.

[13] Signoretto, Marco, Lieven De Lathauwer, and Johan AK Suykens. "Nuclear norms for tensors and their use for convex multilinear estimation." Submitted to Linear Algebra and Its Applications 43 (2010).

[14] Lacroix, Timothée, Nicolas Usunier, and Guillaume Obozinski. "Canonical tensor decomposition for knowledge base completion." International Conference on Machine Learning. PMLR, 2018.

[15] Candes, Emmanuel, and Benjamin Recht. "Exact matrix completion via convex optimization." Communications of the ACM 55.6 (2012): 111-119.

[16] Nickel, Maximilian, Volker Tresp, and Hans-Peter Kriegel. "A three-way model for collective learning on multi-relational data." Icml. 2011.

[17] Zhang, Jianguang, Yahong Han, and Jianmin Jiang. "Tucker decomposition-based tensor learning for human action recognition." Multimedia Systems 22.3 (2016): 343-353.

[18] Ruffinelli, Daniel, Samuel Broscheit, and Rainer Gemulla. "You can teach an old dog new tricks! on training knowledge graph embeddings." ICLR 2020: Eighth International Conference on Learning Representations 4.26-5.1 (2020): 1-12.

[19] Lacroix, Timothée, Nicolas Usunier, and Guillaume Obozinski. "Canonical tensor decomposition for knowledge base completion." International Conference on Machine Learning. PMLR, 2018.

[20] Zheng, Zibin, et al. "QoS-aware web service recommendation by collaborative filtering." IEEE Transactions on services computing 4.2 (2010): 140-152.

[21] Chai, Sheng. "Tensor Decomposition Method Applied to Recommendation Systems." Ph.D. thesis, Queen's University, 2021

[22] Feng, Wenying, et al. "Mining network data for intrusion detection through combining SVMs with ant colony networks." Future Generation Computer Systems 37 (2014): 127-140.

[23] He, Ben, Jimmy Xiangji Huang, and Xiaofeng Zhou. "Modeling term proximity for probabilistic information retrieval models." Information Sciences 181.14 (2011): 3017-3031.

Authorized licensed use limited to: Queen's University. Downloaded on May 31,2023 at 18:30:34 UTC from IEEE Xplore. Restrictions apply.