GSTA: Pedestrian trajectory prediction based on global spatio-temporal association of graph attention network

Citation for published version (APA):

Document license:
TAVERNE

DOI:
10.1016/j.patrec.2022.06.011

Document status and date:
Published: 01/08/2022

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain.
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
GSTA: Pedestrian trajectory prediction based on global spatio-temporal association of graph attention network

Wei Kong, Yun Liu, Hui Li, Chuanxu Wang, Ye Tao, Xiangzhen Kong

A School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China
B Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven 5600MB, the Netherlands

ABSTRACT

Most encoder-decoder structure based predictions models usually predict trajectory according to the position and historical movement of nearby pedestrians. Their input range (receptive field) is small. They often ignore some specific information such as the speed and direction of pedestrians’ movement or the temporal attention. This leads to detailed pedestrian interaction that cannot be obtained. Therefore, we propose a novel spatio-temporal graph attention network (GAT) called GSTA. In the spatial domain, GSTA obtains complex interaction by spatial attention (SA) based on multi-feature fusion, and expands the receptive field through feature updating mechanism (FUM). In the temporal domain, we design temporal attention module (TAM) and feature selection module (FSM). TAM is used to discover the internal relationship of historical trajectory and solve the problem that temporal attention is averaged. FSM overcomes the adverse effect of small temporal perceptual range and reasonably controls the flow of feature information. Experimental results on 5 commonly used pedestrian trajectory datasets show that the prediction accuracy of our proposed model is further improved.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

In the field of autonomous driving [1], object tracking [2] and human-robot interaction [3], the research on pedestrian trajectory prediction has a broad application prospect. Due to the complexity and randomness of pedestrian interaction, pedestrian trajectory is difficult to predict. The traditional methods [4,5] take into consideration specific interactions among pedestrians, and don’t perform well in complex scenes. Recurrent Neural Network (RNN) and its variants show strong modeling ability, and are widely applied in pedestrian trajectory prediction [6–8]. Although RNN-based models have achieved better results than previous studies, these are aggregation methods (such as pooling) to model pedestrian interactions. They don’t deal with spatial context alone, but need to encode adjacent information with additional structure.

Graph Convolutional Network is another widely applied model in pedestrian trajectory prediction. It is more intuitive and effective than aggregation methods. Due to different importance of pedestrians in trajectory prediction, attention mechanism is more helpful to encode the relative influence and potential interaction among pedestrians. Some methods model pedestrians in the scene as a fully connected graph, and introduce a flexible graph attention mechanism to get dynamic pedestrian interactions[9,10]. However, these methods only consider close pedestrians and obtain local information, leading to unreasonable attention allocation and small perception range of the network.

As shown in Fig. 1(a), first, for pedestrian \(i\), if the influence on his trajectory is judged according to distance \(L\), speed \(V\), and direction \(\theta\) respectively, pedestrian \(j\), runner \(k\) and pedestrian \(l\) have different importance. Therefore, spatial attention is not only af-
ected by distance. Moreover, when fusing features for pedestrian $i$, pedestrian $j$ and pedestrian $l$ are selected according to distance, while features of the runner $k$ are ignored, which makes the input range of the network smaller. Second, the temporal attention is often ignored in research [9–11], which makes little difference in the attention of pedestrians at different time steps. In Fig. 1(a), due to different historical trajectories of each pedestrian at $t_1$, $t_2$, $t_3$, influence weight of each pedestrian on the pedestrian $i$ will also change. Therefore, the motion state of the target pedestrian at different time $t$ has different effects on future trajectory prediction. Third, the temporal receptive field of the network is small. In the temporal domain, pedestrian trajectory prediction based on LSTM only depends on the hidden state of the previous moment, and can't be processed in parallel as Convolutional Neural Networks (CNN), as shown the missing connections in Fig. 1(b). Running time of the model is long and perception range is narrow. Our contributions are summarized as follows:

- We propose a novel model called GSTA. In the spatial domain, we design SA and FUM in GSTA. SA integrates multi-feature information and helps the target pedestrians capture subtle changes in the motion state of the surrounding pedestrians. FUM increases the spatial receptive field and enhances the association of global features.

- In the temporal domain, we present TAM and FSM. TAM obtains the internal correlation of historical trajectories and solves the problem that temporal attention is averaged. FSM increases temporal receptive range and controls the flow of feature information.

- GSTA has superior performance in improving prediction accuracy and high computational efficiency in reasoning process comparing with the current state-of-the-art methods.

2. Related work

*Human-human interactions.* The modeling method of human-human interactions has gone through the stages of social force model, multimodel method, hybrid estimation and pattern-based method. The pattern-based method fits different functions (such as neural network) from data to learn human interaction, which improves the flexibility of the model. RNN for sequence learning has become the mainstream of pedestrian interaction modeling, and it is easy to combine with other structures. For example, RNN and CNN jointly model spatial relationship [12] to capture interactions among agents. The population descriptor based on social perception LSTM is combined with Gaussian Process [13] to predict complete distribution of future trajectory. However, RNN is prone to gradient vanishing and exploding in the training of long sequence. For this reason, we use LSTM and combine residual connections to solve the problem of long sequence dependence.

*Trajectory prediction based on graph model.* Although RNN has powerful sequence modeling ability, it lacks intuitive spatio-temporal structure. Graph structure is a natural way to represent pedestrian interaction. Graphs usually represent pedestrians as nodes, their interactions as edges, and combine them with deep sequence models (such as LSTM) [14,15]. Some methods are developed on this basis. Zhang et al. [16] dynamically constructed a directed social graph based on position and movement direction. Liang et al. [17] proposed a scene semantic graph to model the interaction between pedestrian and scene. However, these methods only capture local pedestrian interaction based on distance, and can't reflect the global pedestrian relationship. In this paper, a fully connected spatio-temporal graph is established, and the global feature association is enhanced by FUM.

*Trajectory prediction based on attention mechanism.* GAT-based trajectory prediction methods [9,10,18] capture the features of the target pedestrian through neighborhood weighted aggregation. But the calculation of attention is simple, ignoring deep movement features, such as the fusion of distance, speed and direction. Although some methods try to fuse the distance and speed features of pedestrians [16,19], the accuracy of trajectory prediction is not satisfactory. In the temporal domain, the attention of pedestrian movement at different time steps is averaged. SA in GSTA breaks through the limitation of only using location data and achieves

---

*Fig. 2.* The overall framework of GSTA.
more realistic attention distribution. In the temporal domain, GSTA allocates more appropriate attention to pedestrians at each time step by TAM.

**CNN for pedestrian trajectory prediction.** The RNN-based prediction models [6,8,20] only depend on the output of previous time, ignoring the influence on trajectory prediction at other time steps. CNN can achieve parallel processing and extract rich context information. Yi et al. [21] used a large receptive field to simulate pedestrian behavior. Yagi et al. [22] developed a deep neural network to predict pedestrian location. However, only using CNN to concentrate the features of nearby pedestrians will lose some motion information, which limits the prediction accuracy. To improve the temporal receptive field and capture significant pedestrian features, FSM for time reasoning is designed in GSTA. Experiments show that this design further improves prediction performance.

### 3. Methods

GSTA is an encoder-decoder structure. The overall framework of GSTA is shown in Fig. 2. SA, FUM, TAM, FSM are our innovations.

**3.1. Problem definition**

Assuming that the position of a pedestrian $i$ at time $t$ is $(x_i^t, y_i^t)$, the trajectories of $N$ pedestrians can be expressed as $p_i^t = \{ (x_i^t, y_i^t) \mid t = 1, \ldots, T_m, i = 1, \ldots, N \}$. Trajectory prediction is aimed to predict the future trajectory $p_i^t$ of $N$ pedestrians during time steps $T_{m+1}, \ldots, T_{end}$. $p_i^t = \{ (x_i^t, y_i^t) \mid t = T_{m+1}, \ldots, T_{end}, i = 1, \ldots, N \}$.

**3.2. Encoding for a single pedestrian**

LSTM has been proved to be able to describe pedestrian motion patterns [6,7,9]. To enhance the current information transmission, we add residual connection to LSTM to form TS-LSTM. Firstly, displacement of all pedestrians at the current time relative to the previous time is calculated and embedded as a fixed vector, as shown in Eq. (1), $\varphi$ is an embedding function, $W_h$ is an embedding weight.

\[
(\Delta x_i^t, \Delta y_i^t) = (x_i^{t+1} - x_i^t, y_i^{t+1} - y_i^t) \\
W_i^t = \varphi(\Delta x_i^t, \Delta y_i^t, W_h)
\] (1)

Next, the hidden state of TS-LSTM is calculated in Eq. (2). $W_h$ is a shared weight of TS-LSTM.

\[h_i^t = \text{LSTM}(h_i^{t-1}, v_i^t, W_h) + v_i^t\] (2)

**3.3. FUM based on GAT**

To expand the spatial receptive field of our network, feature updating (FU) is added to GAT to form FUM. FU is to get the global features by calculating intimacy for all pedestrians in graph $G$. The intimacy function $d(h_i^t, h_j^t) = e^{d(h_i^t, h_j^t)}$, $h_i^t$ is reshaped into $h_i^\|$. In Eq. (3), $s(\cdot)$ is a display function for calculating features of adjacent pedestrians. $c_i^t$ is reshaped into $c_i^\|$.

\[
M_i^t = \frac{d(h_i^t, h_j^t)}{\sum_h d(h_i^t, h_j^t)} \\
s(h_j^t) = w_i h_j^t, c_i^t = M_i^t s(h_j^t) \\
z_i^t = w_c c_i^t + h_i^\|
\] (3)

SA. It is not enough to learn pedestrian interaction only based on distance $A_{dir} \in R^{N \times N}$. Since the dataset is sampled every 0.4 s,
the calculation of relative speed $A_{spd} \in \mathbb{R}^{N \times N}$ is converted into relative distance divided by sampling time, and the direction $A_{dir} \in \mathbb{R}^{N \times N}$ is calculated by cosine similarity between pedestrians. $A_e$ is updated in Eq. (4). $T$ represents transposition, $a \in \mathbb{R}^{D'}$ is a weight vector of single layer perceptron, $W \in \mathbb{R}^{D' \times D'}$ is a shared weight to realize linear transformation. $D$ is the feature dimension of $h^i$, $D'$ is the output dimension. $\|$ is a concatenation operation. $j$ represents neighbors of pedestrian $i$.

\[
\alpha_{t}^{ij} = \text{Softmax} \left( \text{LeakyReLU} \left( a^T [W^i h^i_t \| W^j h^j_t] \right) \right)
\]
\[
A_{dir} = \{ \alpha_{t}^{ij} \mid \forall ij \in \{1, \ldots, N\}, t = 1, \ldots, T_m \}
\]
\[
A_e = \text{Softmax} \left( A_{dir} \otimes (A_{spd} + A_{dir}) \right)
\]

Output of FUM. Combining $A_e$ and $Z$ (from Eq. (3)), the output of FUM is the result of graph convolution in Eq. (5). Where, $Z^{(l)} \in \mathbb{R}^{N \times D_l}$ is the feature matrix of $N$ pedestrians on layer $l$ ($l = 2$). $W_e \in \mathbb{R}^{D_l \times D_l \times 1}$, $Z'$ in $Z^{(l+1)}$ is concatenated by multi-head attention. The

Fig. 5. Comparison of attention allocation between GSTA and STGAT in different scenes.
Fig. 6. Comparison of predicted trajectories between GSTA and STGAT in different scenes.
number of attention heads is 4. Adding self connection to $A_e$ is $\tilde{A}_e$.
\[
Z_t^{(i-1)} = \sigma(\tilde{A}_e Z_t^{(i)} W_e)
\]  
(5)

3.4. TAM

Given the input $Z$ (from Eq. (5)), the output becomes $R$. In Eq. (6), $Z$ is transformed into three tensors, and $f$ is a linear transformation function.
\[
Q_i = f_0\{z_i^{(1)}|z_i^{(2)}|z_i^{(3)}\}
\]
\[
K_i = f_k\{z_i^{(1)}|z_i^{(2)}|z_i^{(3)}\}
\]
\[
V_i = f_v\{z_i^{(1)}|z_i^{(2)}|z_i^{(3)}\}
\]  
(6)

One head attention of TAM is recorded as $head_j$ in Eq. (7). $s_j$ is a scaled factor to adjust the output to a reasonable range.
\[
head_j = \text{Softmax}(Q_j K_j^T) / s_j V_j
\]  
(7)

The multi-head attention is shown in Eq. (8). Where, $h_{:.num} = 8$ is the number of attention heads.
\[
R_i = f_{\cdot}(\sum_{j=1}^{h_{:.num}} head_j)^{h_{:.num}}
\]  
(8)

3.5. Output of encoder

Feature fusion. LSTM-based methods can share hidden state [6,11] to achieve spatio-temporal fusion, so TG-LSTM is proposed. TG-LSTM also adds the residual connections to LSTM. In Eq. (9), $r_i^T$ comes from Eq. (8) and $W_e$ is the weight of TG-LSTM.
\[
g_{i}^{T} = \text{LSTM}(g_{i-1}^{T}, r_{i}^{T}, W_{e}) + r_{i}^{T}
\]  
(9)

Output of encoder. The trajectory change of the target pedestrian not only comes from the interactions of surrounding pedestrians, but also depends on the influence of target pedestrians themselves. Therefore, we concatenate $h_{i}$ (from Eq. (2)) and $g_{i}^{T}$ to complete encoding, as shown in Eq. (10). $\delta$ is a multilayer perceptron.
\[
c_{i}^{T} = \delta(h_{i})\|\delta(g_{i}^{T})
\]
\[
C_{0} = \{c_{i}^{T} | i = 1, \ldots, N, t = 1, \ldots, T_m\}
\]  
(10)

3.6. FSM

Convolution layers. In Fig. 3(a), FSM has three convolution layers, and the convolution kernel is $3 \times 3$. Observing changes of red lines, the receptive field becomes larger and larger with deepening of convolution layers.

Feature selection. A gated mechanism composed of two activation functions is designed in Fig. 3(b). The output is shown in Eq. (11). $b$ is the bias, $C_3$ is the output of convolution layers, and $\odot$ is the hadamard product.
\[
O = \tanh(W_o C_3 + b_a) \odot \sigma(W_o C_3 + b_\sigma) + C_3
\]  
(11)

3.7. P-LSTM for trajectory prediction

P-LSTM is similar to TS-LSTM/TG-LSTM. In the training process, random noise $U$ which obeys the standard normal distribution $N(0, 1)$ is sampled. $U$ and $O$ are concatenated as the input of P-LSTM in Eq. (12).
\[
e_{i}^{\tilde{m}} = O_{i}^{\tilde{m}} || u
\]  
(12)

The hidden state of P-LSTM is $e_{i}^{\text{m}+1}$ in Eq. (13). $e_{i}^{\text{m}}$ is the initial hidden state, $v_{i}^{\text{m}}$ (from Eq. (1)) represents the initial input, and $W_e$ is an updatable weight of P-LSTM.
\[
e_{i}^{\text{m}+1} = \text{LSTM} (e_{i}^{\text{m}}, v_{i}^{\text{m}}, W_e) + e_{i}^{\text{m}}
\]
\[
(\Delta y_{i}^{\text{m}+1}, \Delta y_{i}^{\text{m}+1}) = \delta_{x}(e_{i}^{\text{m}+1})
\]  
(13)

3.8. Definition of loss function

We use diversity loss strategy proposed by social Gan [11]. $k$ results can be generated in one training, $k = 20$. We take the minimum value as the loss, as shown in Eq. (15). $Y_i$ is the actual trajectory, $\hat{Y}_i$ is the predicted trajectory.
\[
\text{Loss} = \min_{k} \| Y_i - \hat{Y}_i \|_2
\]  
(15)

4. Experiments and results analysis

4.1. Experimental settings

We verify the experimental results on two open datasets, ETH and UCY. The ETH dataset contains two scenes named ETH and HOTEL, and the UCY dataset contains three scenes named UNIV, ZARA1 and ZARA2. The evaluation metrics are ADE (average displacement error) in Eq. (16) and FDE (final displacement error) in Eq. (17). GSTA is implemented in PyTorch framework, and its corresponding version is 1.2. Two NVIDIA GeForce GTX-1080 GPUs are used for training. Our model is optimized with Adam and the batch size is 64. The size of hidden state and output of TSG-LSTM is 32 dimensions. Embedded vector $v_i$ is 32 dimensions. The scaled factor $s_f=0.5$. The size of noise $u$ is set to 16 dimensions. During training, the method leave-one-out is adopted. The observed trajectory is 3.2 s (8 time steps), and the predicted trajectory is 4.8 s (12 time steps).
\[
\text{ADE} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_{end}} \| \hat{Y}_{i}^{t} - Y_{i}^{t} \|_2}{N \times T_{end}}
\]  
(16)
\[
\text{FDE} = \frac{\sum_{i=1}^{N} \| \hat{Y}_{i}^{T_{end}} - Y_{i}^{T_{end}} \|_2}{N}
\]  
(17)
4.2. Ablation study

We compare six methods in Table 1. The effectiveness of SA, FUM, TAM and FSM in GSTA is verified. Top-1, Top-2, Top-3 results are shown in red, green, and blue (the same below). Res means combining the LSTM with residual connections.

Experimental results of the above six methods are shown in Table 1, we found that:

1) Direction and speed information in SA are beneficial to improve model performance. The fusion of multiple motion information for pedestrians can better determine the motion state of pedestrians and achieve more accurate encoding.

2) The combination of SA and TAM makes the overall performance of the model better than STGAT, which indicates that SA and TAM provide better spatial and temporal modeling capabilities.

3) The good experimental results of EGAT are mainly due to the improvement of spatial input range by FUM(FU+GAT) and the enhancement of information transmission by residual connections. FUM realizes feature sharing among pedestrians, and residual connections enable feature information effectively associated.

4) Compared with STGAT and EGAT, our model GSTA has better experimental results, which is mainly affected by FSM in the temporal domain. For FSM, convolution layers improve the temporal perceptual range, and our model can obtain more abundant perceptual information. The feature selection can associate and filter features at different time steps.

4.3. Comparison with the state-of-the-art

**Inference time.** In Table 3, our method is high computational efficiency. This is because we only use visual information and don’t need to detect and track pedestrians in the scene. Because our model uses recursive network LSTM for temporal reasoning, it is slower than Social-STGCNN in reasoning speed. However, compared with STGAT, our model is still better. This is mainly because our model increases the receptive field and improves the efficiency of data parallel processing.

**Evaluation metrics analysis.** The experimental results show that GSTA outperforms STGAT in all scenes in Table 4, and two AVGs are reduced by 14% and 7.3% respectively. In UNIV, the reason why GSTA is not better than SOTA is that the high-density population involves more pedestrian interaction, forcing the target pedestrian to make decisions among different options, which makes the prediction more challenging. In ZARA1, the trajectory of pedestrians is often affected by the surrounding pedestrians and obstacles, which may change or limit human activities, resulting in the model being unable to capture more social interactions.

Compared with SGCN, GSTA performs best on ETH and HOTEL datasets with less pedestrian interaction. But on UNIV and ZARA datasets with dense pedestrians, GSTA performs poorly. This is because the model with simple pedestrian interaction, the change of graph structure is small, and the fixed graph structure has little influence on GSTA, which makes the prediction performance of GSTA better. When pedestrian interaction becomes complex and changeable, the fixed graph structure cannot be adjusted adaptively with the change of scene, so the performance of GSTA decreases on UNIV and ZARA datasets. Therefore, Pedestrian trajectory prediction based on adaptive directed graph lays a foundation for future research.

However, compared with SGCN, our model also has many advantages. Firstly, in terms of inference time, GSTA is faster than SGCN, as shown in Table 3. Secondly, in the simple scenes ETH and HOTEL, our model performs best, while SGCN performs poorly. This shows that GSTA is suitable for the application of simple models, while SGCN is easy to over fit. Thirdly, to verify the influence of flexible graph structure on our model and prove that our model has advantages over SGCN, this paper constructs an adaptive spatial graph based on GSTA and proposes the model GSTA+. The ablation experiment in Table 2 is on ZARA2 dataset, and the predicted length is 8 time steps. The data in the table shows that the performance of GSTA+ is better than SGCN, which proves that GSTA+ has development potential.

4.4. Analysis of experimental results

**Training efficiency.** In Fig. 4, the fitting speed of GSTA is faster than STGAT, and GSTA is more stable with increased epochs. Although STGAT can adapt to the sample, its fitting ability to the sample is not robust.

**Variability of attention.** Pedestrians walking in the opposite direction (Fig. 5(a) and (b)) or in front (Fig. 5(c) and (d)) receive more attention than those walking in parallel or behind in GSTA. STGAT ignores the influence of direction on pedestrian trajectory. In Fig. 5(e), for a stationary target pedestrian, GSTA pays less attention to distant pedestrians and pedestrians without intention conflict, while STGAT allocates more attention to them. In Fig. 5(f)-(h), for the moving target pedestrian, we should ignore the long-distance stationary pedestrian and pay more attention to the reverse and fast pedestrian. But STGAT shows the opposite result. These successful results show that compared with STGAT, GSTA learns more reasonable attention weight and better captures the global information of surrounding pedestrians.
Diversity of predicted trajectory. In Fig. 6(a) and (b), the accuracy of STGAT is worse than GSTA. For group movement, GSTA can better infer the change of pedestrian trajectory, as shown in Fig. 6(c) and (d). When pedestrians move nonlinear(such as turning), the prediction performance of STGAT is worse than GSTA. For example, in Fig. 6(f)(3), STGAT predicts that the woman in black T-shirt will be waiting in place. But GSTA infers that she is about to cross the road. For the nonlinear motion of multiple pedestrians in Fig. 6(g), we can see that GSTA is more accuracy than STGAT. In addition, in Fig. 6(b)(1), (c)(1) and (h)(2), GSTA accurately judges stationary pedestrians (the trajectory is represented by points) according to the speed, while STGAT regards the stationary pedestrian as a moving pedestrian. Compared with STGAT, GSTA can receive a wide range of feature information and capture the subtle changes of pedestrian interaction.

Problems and research direction. When there are many pedestrians in a scene, as shown in Fig. 5(i), the attention distribution is easy to be averaged. Because the feature difference between pedestrians decreases with increased number of pedestrians. When the pedestrian turns sharply in the scene, as shown in Fig. 6(h), the model can’t judge the pedestrian's motion intention in time. We attribute these failure cases to the lack of additional auxiliary information, such as scene information, pedestrian social attribute information and so on.

5. Conclusion

In this paper, a novel model GSTA for pedestrian trajectory prediction is proposed, which aims to solve the problems of single spatial interaction, small spatio-temporal perception range and no difference in temporal attention. In the spatial domain, spatial attention (SA) and feature updating mechanism (FUM) are proposed. SA fuses different motion features to enhance pedestrian interaction. FUM increases the spatial receptive field to associate global features. In the temporal domain, temporal attention module (TAM) and feature selection module (FSM) are designed. TAM can motivate the network to adjust the temporal weight of pedestrian motion state and further simulate the real scene. FSM improves the temporal receptive field and selects important pedestrian features to improve the prediction performance. Experiments show that GSTA can allocate reasonable attention to pedestrians, improve the perception range of the model, and predict more reliable trajectory according to the changes of different scenes. Although this model has achieved good prediction performance, there are some problems, such as attention averaging in dense scenes, the fixed graph structure and inaccurate grasp of pedestrian motion intention. These are the contents to be studied in the future.

Declaration of Competing Interest

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

Acknowledgements

All authors are very grateful to the National Natural Science Foundation of China (Nos. 61702295, 61672305) and the Shandong Province Colleges and Universities Young Talents Initiation Program (No. 2019KJN047) for research support.

References