Using Deep Learning Methods to Analyze User Behavior in CNGrid

Xiaodong Wang

中国科学院
计算机网络信息中心
Computer Network Information Center, Chinese Academy of Sciences

Catalogue

- Instruction
- Background
- Preprocessing Module
- LSTM Language Model
- Attention Mechanism
- User Behavior Mapping
- Anomaly Detection of Single User
- The Experiment and Analysis
- Conclusion

Introduction

- User behavior is complex and variable
- There are many users in a supercomputing environment
  - Each user behavior has its own characteristics
- User behavior includes user operational characteristics
- It needs to be analyzed using some automated detection methods

- An unsupervised learning algorithm based on deep learning is introduced to analyze user behavior in network environment
  - Data preprocessing
  - Use of LSTM language models
  - Attention mechanism is used to improve accuracy
- Use Case
  - The user behavior map can be obtained by using this model
  - This model can be used for single-user exception detection

Background

- The early stage of the work
  - Thinking based on log analysis anomaly detection model
  - Preprocessing of log data modeling
  - Heuristics for log type operation lists

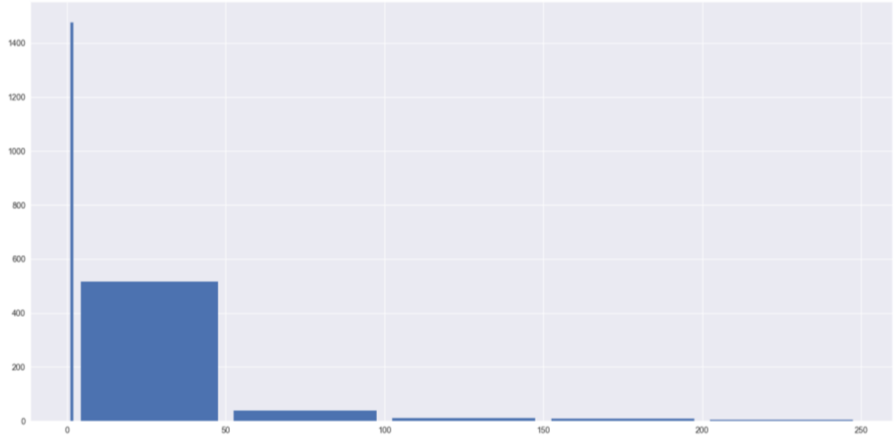
- Definition1： For user operation Op_i $(i \in 1,2,\ldots,n)$， Which n express The operation number of user . $UserBehavior = \{Op_{i1}, Op_{i2}, \ldots, Op_{im}\}$ Represents a user action.

Operation type	Operation Type interpretation	Operation type number
end	The end tag	0
SCE_CMD_MREMOTE	FS命令执行	1
SCE_BJOBS_ENC	Job list	2
SCE_JOB_LISTRES_ENC	List resources	3
SCE_JOB_SUBMIT	Submit job	4
SCE_CMD_REMOTE	FS Command execution	5
Start	The start tag	6

An ordered connection between different types of user operations represents a session operation

Preprocessing Module

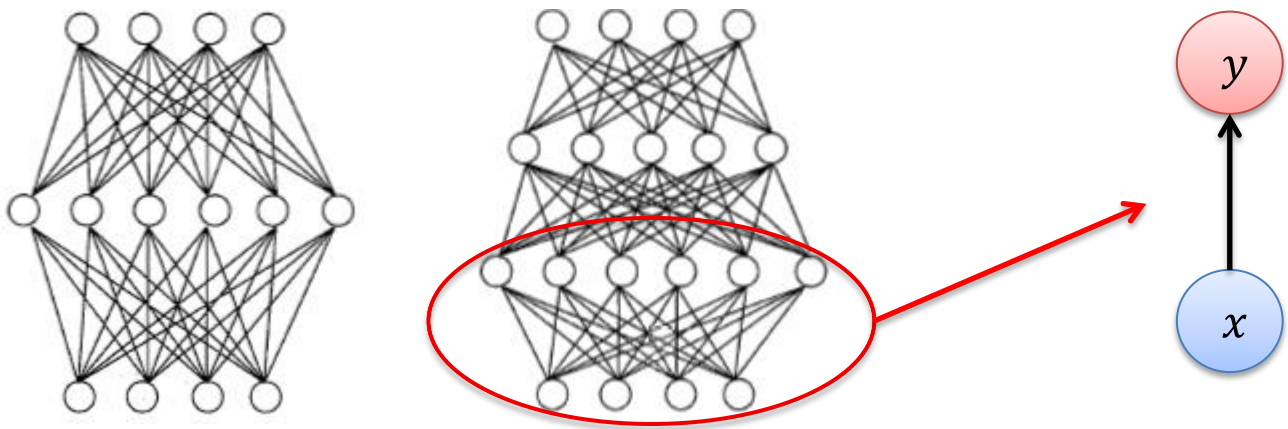
- A trapezoid diagram of the number of operations under different sessions



- Filter small number of operations reason
  - The number of one operation sessions is large
  - A single action does not make much sense for an ordered list analysis of user actions
- Filter large number of operations reason
  - Sessions larger than 100 appear very infrequently

LSTM Language Model

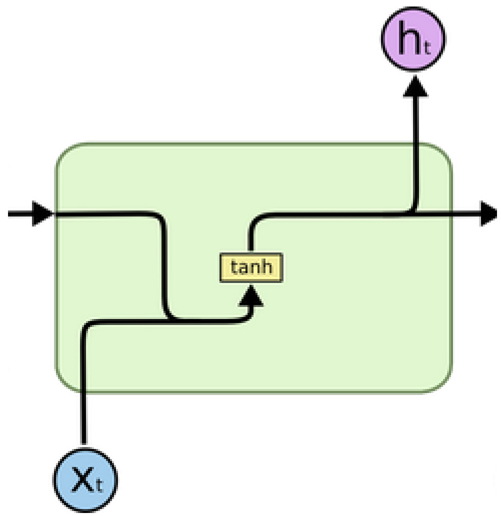
- RNN basic structure



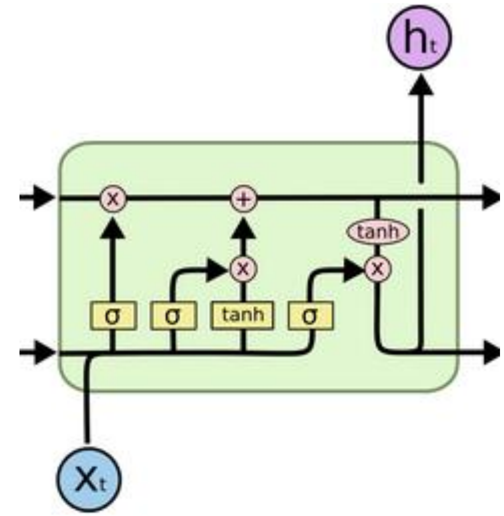
$y = f(Wx + b)$

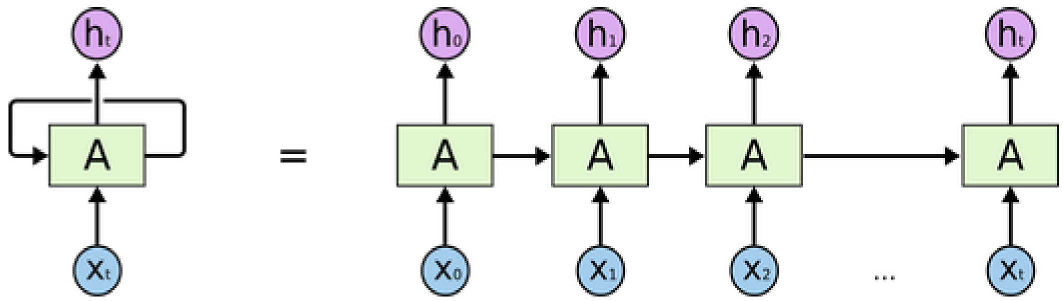
- Sequential data model
  - Natural language problems, x_1 for the first word, x_2 for the second word, and so on
  - Time series, take the daily stock price
  - User behavior analysis, the type of each user action

LSTM Language Model



RNN to LSTM

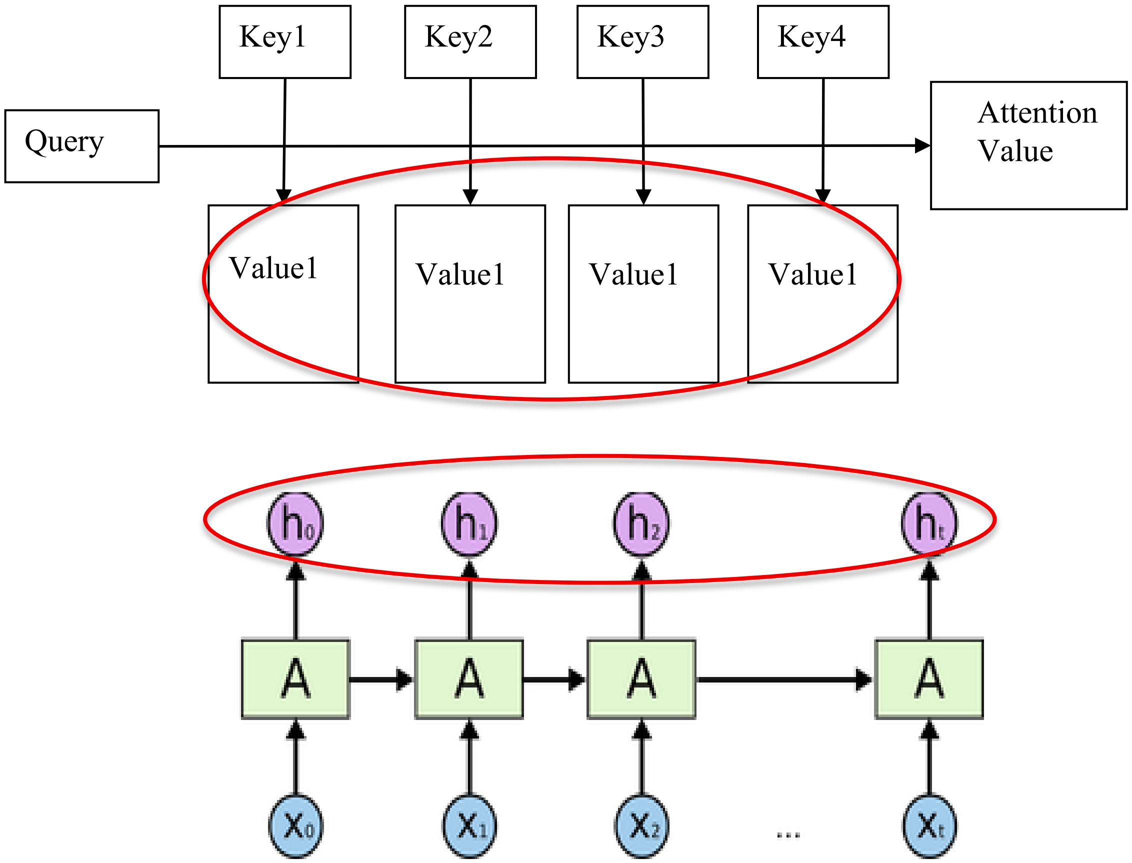




Source of the above picture： http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Attention Mechanism

- The nature of the Attention function can be described as a mapping from a query to a set of key-value pairs



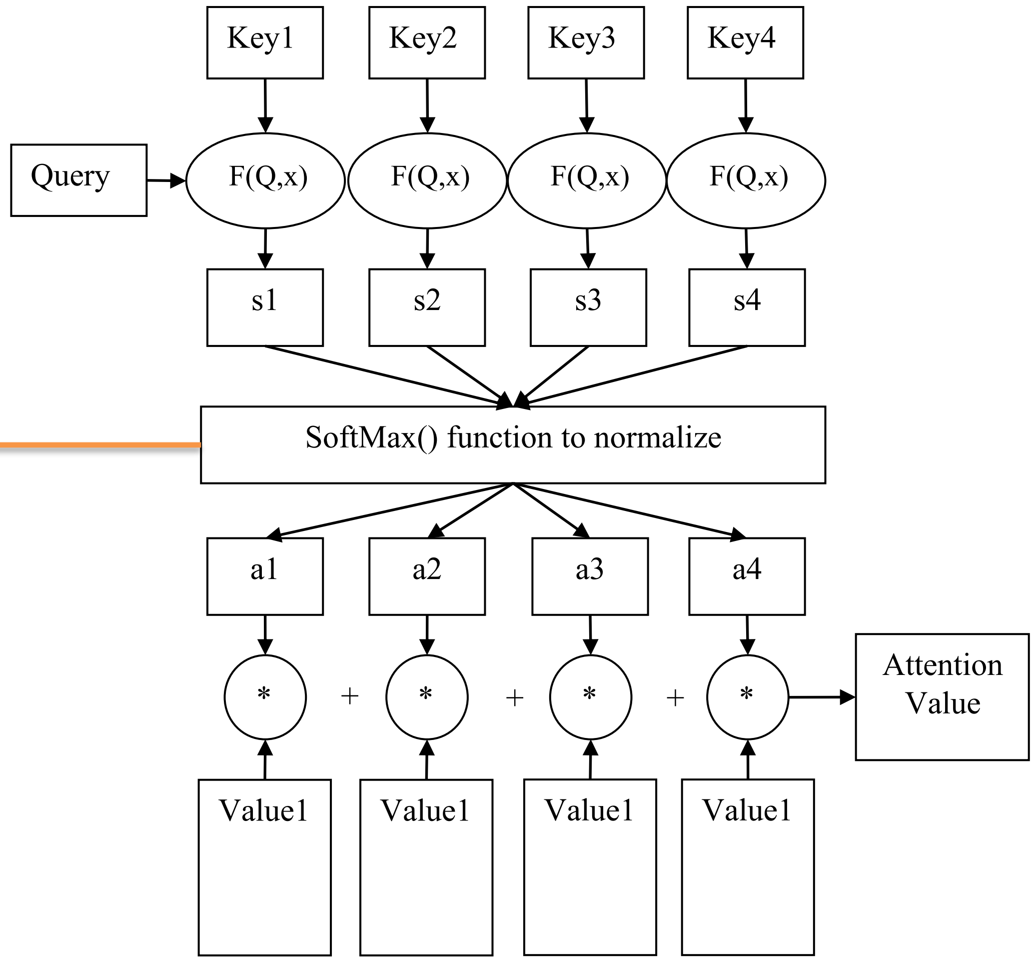
Attention Mechanism

- Dot product, splicing, perceptron

$$f(Q,k) = Q^T K$$
$$f(Q,k) = W_a[Q;K]$$
$$f(Q,k) = V tanh(WQ + UK)$$

$$a = softmax(f(Q,K))$$

$$Attention = \sum aV$$



Attention Mechanism

- The value matrix

$$V_{(t)} = \begin{bmatrix} h_1 \\ \vdots \\ h_t \end{bmatrix} \in R^{t \times L_h}$$

- h_i : hidden state
- t : time steps
- L_h : output dimension at this time of hidden
- The key matrix

$$K_{(t)} = \tanh\left(V_{(t)} \times W^{a}\right) \in R^{t \times L_a}$$

- The weight coefficient:

$$d_{(t)} = softmax\left(\left[K_{(t)} \otimes q_{(t)}\right]^{T}\right) \in R^{t \times t}$$

$\otimes$ will be described at next page.

Attention Mechanism

- Attention result:

$$\mathrm{a}_{(\mathrm{t})} = d_{(t)} V_{(t)}$$

- The notation $\otimes$ is calculated as follows:

$$K_{(t)} \otimes q_{(t)} = K_{(t)} \times W^{q} \cdot M_{triu}$$

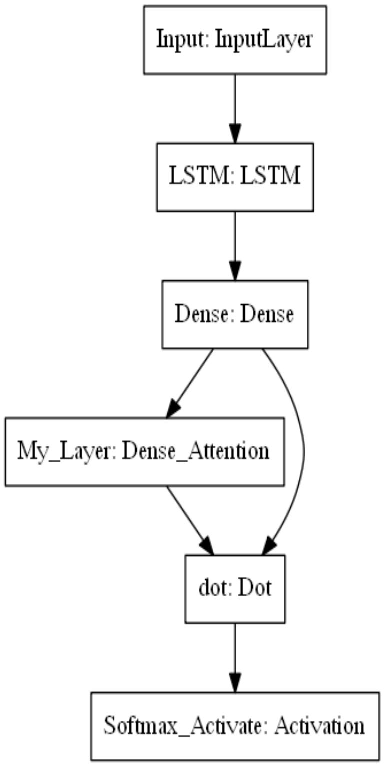
- M_{triu}: a non-zero value of 1 all upper triangular matrix

Thus ensure the $K_{(t)} \otimes q_{(t)}$ is an upper triangular matrix. After transposing, the time step is transposed to the last dimension and the time step dimension is normalized by the activation function softmax().

Attention Mechanism

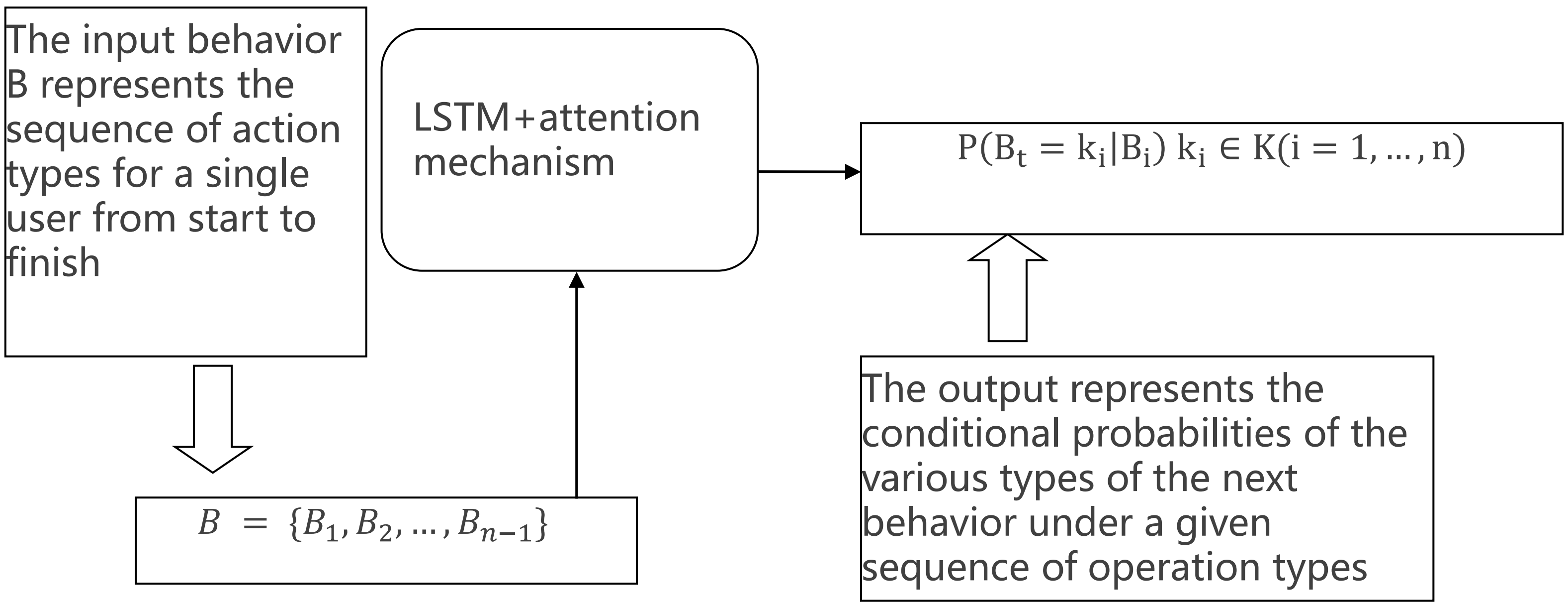
- Final overall structure

Layer Structure	Output the shape of the dimension	Number of parameters (Single LSTM)
InputLayer	(None, 100, 14)	0
LSTM	(None, 100, 128)	73216
Dense	(None, 100, 14)	1806
My_Layer (Dense_Attention)	(None, 100, 100)	7296
dot (Dot)	(None, 100, 14)	0
Softmax_Activate	(None, 100, 14)	0



User Behavior Mapping

- Establishment of prediction model
  - The probability of each type of user behavior occurring the next time is determined by the type sequence



- The user behavior map can be mapped into a multi-fork tree structure
- Each node represents an operation type
- The father-child connection represents the probability of transitioning to the next action

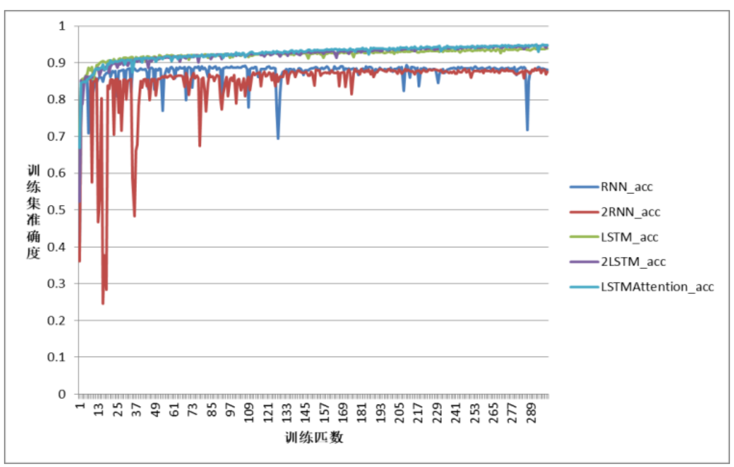
Anomaly Detection of Single User

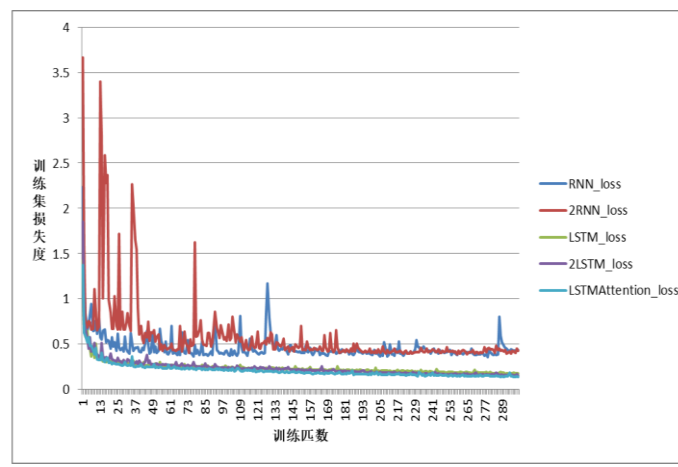
- Mode for single user
  - Determine the selected model (single-layer LSTM+attention mechanism)
  - A sequence of action types for each session of a single user is obtained from historical data
  - Input the model and adjust the model parameters
  - Each user is monitored in real time separately according to the adjusted model

- The maximum probability method is used for anomaly detection
  - The probability of various types of the next action is predicted based on all the previous actions of the user and the model
  - If the first n operation types of the prediction probability of the operation do not match the actual user's operation type, the operation is considered as abnormal behavior

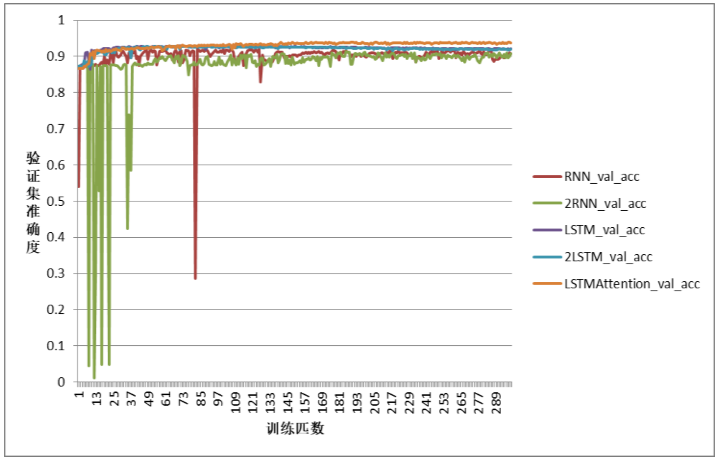
The Experiment and Analysis

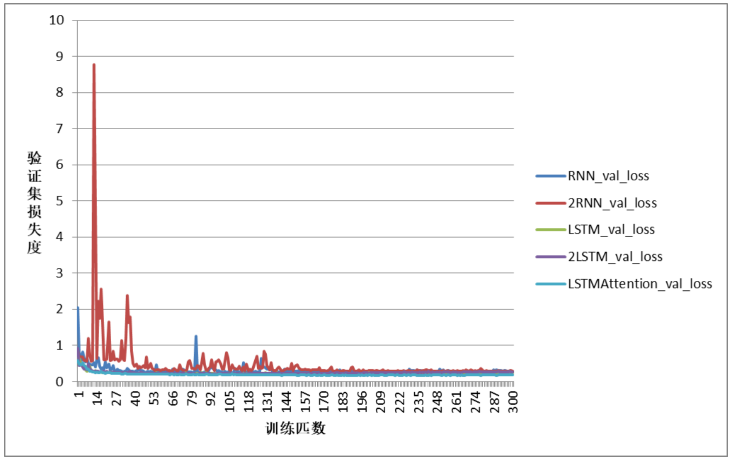
- Data preparation
  - User operation dataset was obtained from CNGrid in the period of 2018/5/9 to 2018/6/12
  - 80% data was used as training set, and 20% data was used as verification set for training and verification
  - Use 100 - dimensional time step input
  - Each type of dimension corresponds to a unique heat encoding of a 14-dimensional vector
  - Single-layer, double-layer RNN, LSTM and LSTM+attention mechanisms were used for training and prediction





The Experiment and Analysis



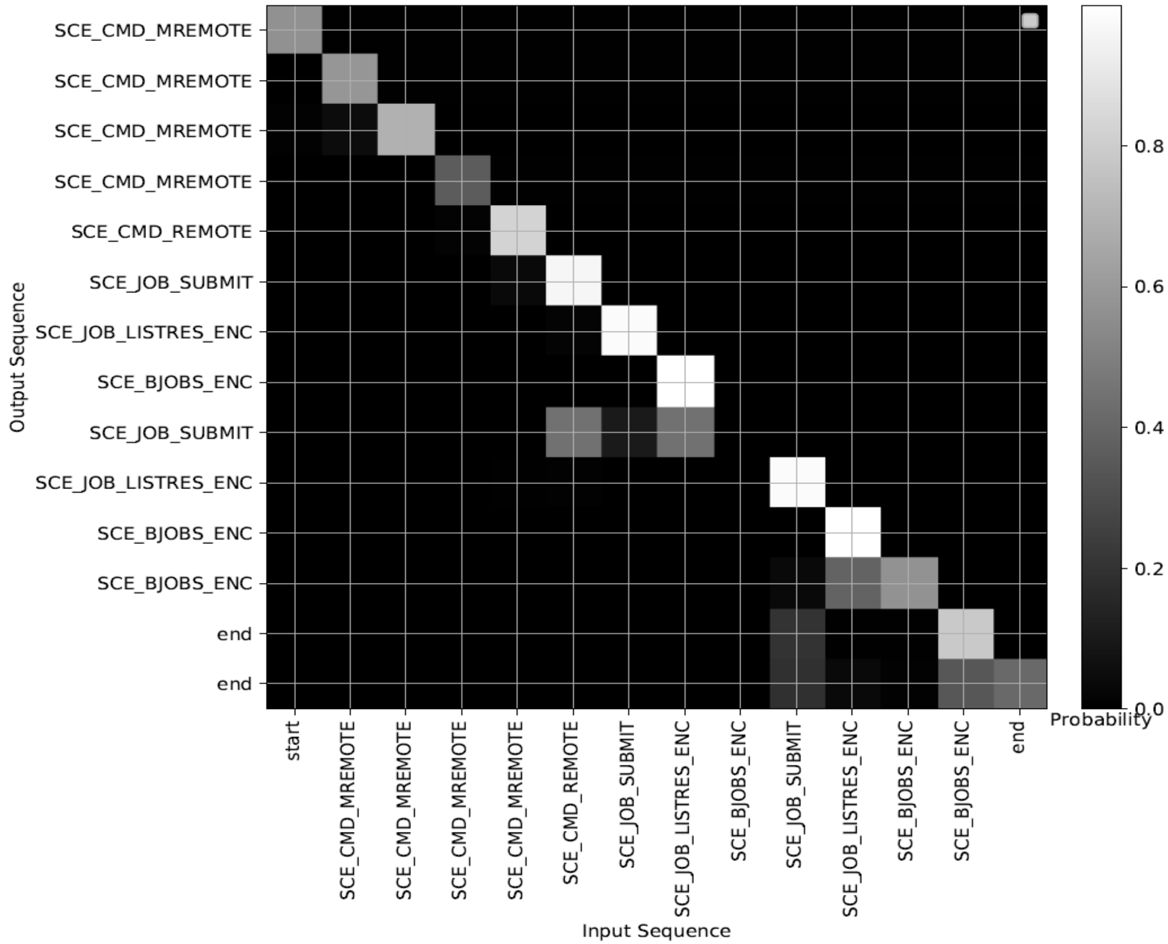


Network type	Highest verification set accuracy	Corresponding batch
Single RNN	92.02%	105
Double RNN	91.29%	185
Single LSTM	92.87%	70
Double LSTM	92.81%	73
Single LSTM + dot attention mechanism	**94.03%**	192

- According to the table above
  - LSTM network is superior to ordinary RNN circulation neural network
  - There is no significant difference in the use of single and double layers between the same networks
  - Attention mechanism can improve the accuracy of deep learning

The Experiment and Analysis

- Attention mechanism figure

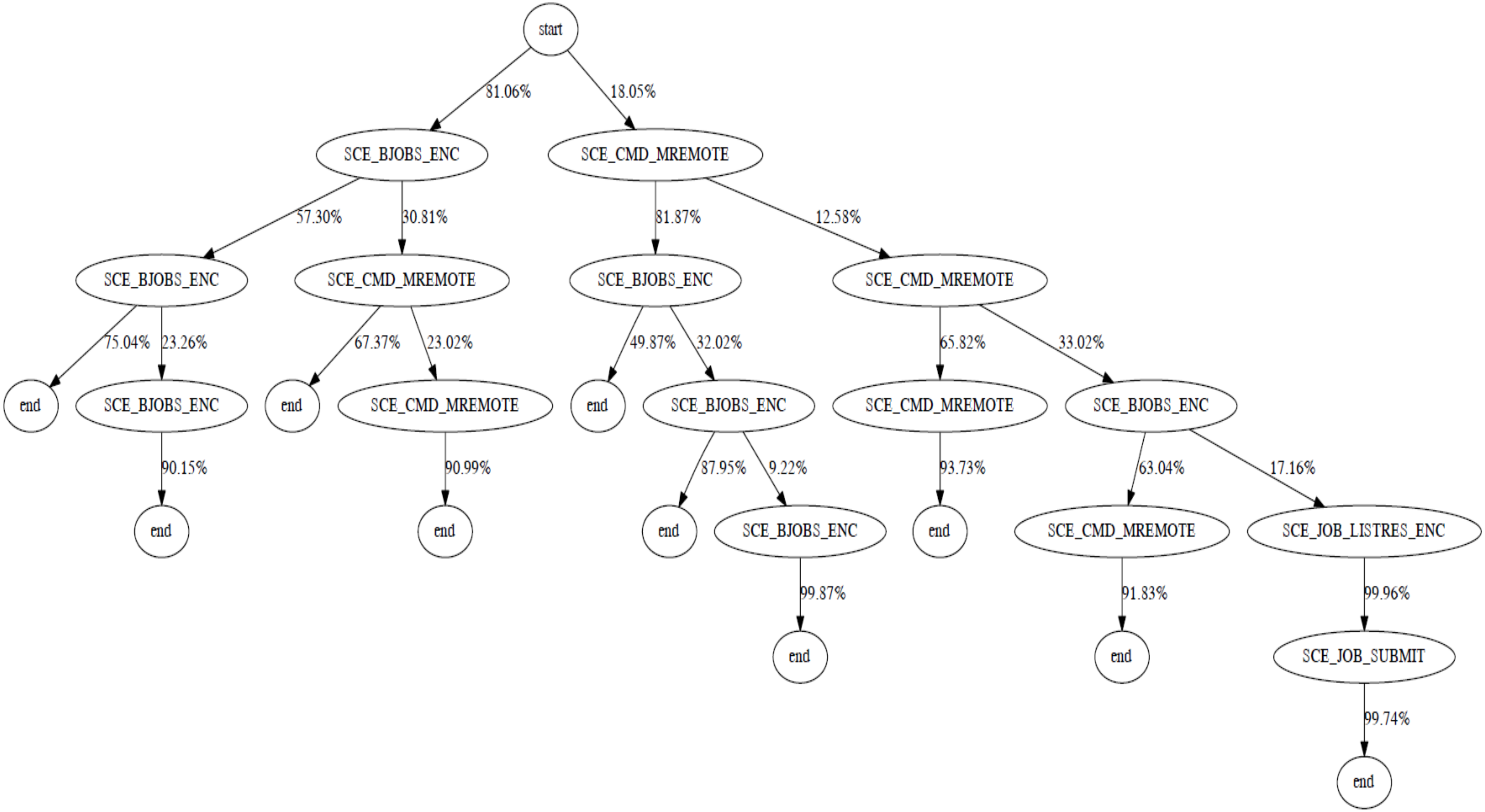


The figure shows the weight of time step obtained by deep learning according to training, and increases the interpretability

The Experiment and Analysis

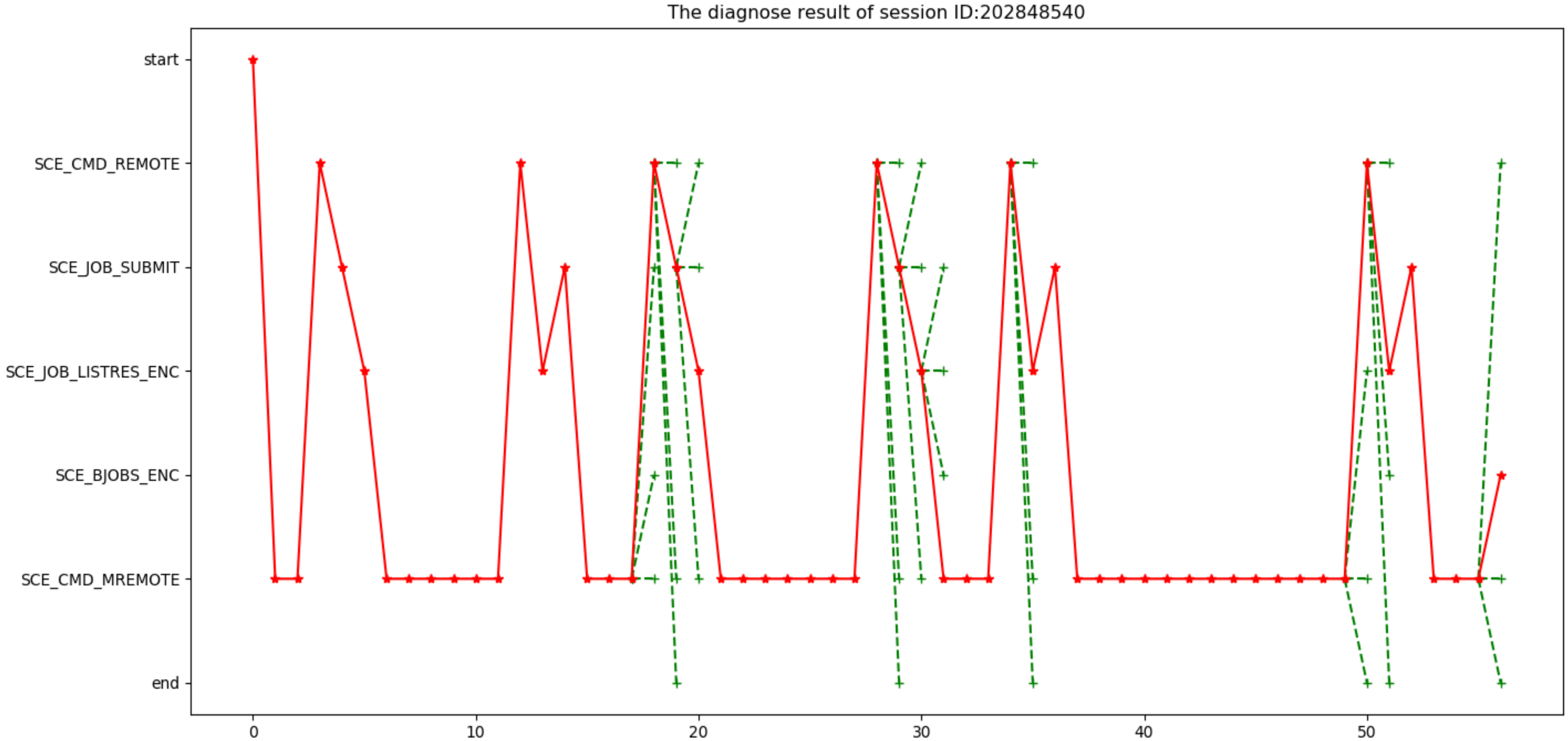
中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

- A case study of single-user behavior mapping
  (User_Example_18/5/9_2018/7/9)



The Experiment and Analysis

- If set n=3, the user (User_Example_18/5/9_2018/7/9) has the following figure of abnormal operation under a session



Conclusion

- Conclusion
  - The user behavior in high performance computing environment is modeled by deep learning cyclic neural network model
  - By comparing the results of different network structure and attention mechanism, the optimal method of constructing user behavior model is obtained
  - The model is used to generate the user behavior map
  - The model is used to detect and analyze user behavior anomalies and provide feedback

- Future Work
  - More refined classification of user operations
  - More kinds of user data from CNGrid

Thank you!

Welcome to visit CNGrid!

中国科学院
计算机网络信息中心
Computer Network Information Center, Chinese Academy of Sciences

Reference

- Reference Articles：
  - Recurrent Neural Network Language Models for Open Vocabulary Event-Level Cyber Anomaly Detection
  - Recurrent Neural Network Attention Mechanisms for Interpretable System Log Anomaly Detection
  - DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning
  - Attention Is All You Need
- Reference Books：
  - http://www.deeplearningbook.org/
- Reference Codes：
  - https://github.com/philipperemy/keras-attention-mechanism
  - https://github.com/datalogue/keras-attention
  - https://github.com/philipperemy/keras-visualize-activations
- Reference Blogs：
  - http://colah.github.io/posts/2015-08-Understanding-LSTMs/