The Influence of Knowledge Worker Salary Satisfaction on Employee Job Performance

Lina Zeng, Capital University of Economics and Business, China*

ABSTRACT

Knowledge workers are quite crucial to every enterprise, so exploring the relationship between their salary satisfaction and job performance is significant. Hence, this work observes their salary satisfaction by identifying the employees’ emotions at the time of salary announcement. The relationship between salary satisfaction and job performance is studied through the obtained satisfaction. First, the convolutional neural network (CNN) model is introduced. Then, it is optimized by adding an attention mechanism to improve the accuracy of the emotion recognition model. Finally, through comparative experiments, the effectiveness of the model proposed and the impact of employee’s salary satisfaction on job performance are verified. The experimental results show that the recognition accuracy of the model is much higher than that of the traditional model. In particular, the recognition accuracy of neutral emotions is as high as 95%. It verifies the effectiveness of the model.

KEYWORDS

Attention Mechanism, Convolutional Neural Network, Emotion Recognition, Job Performance, Salary Satisfaction

INTRODUCTION

Knowledgeable employees refer to those who use knowledge or information to work. In the past, they mainly referred to enterprise managers or executive managers. However, modern knowledgeable employees have been expanded to most white-collar workers. The competition between enterprises; the creation, utilization, and appreciation of knowledge; and the rational allocation of resources are all realized by knowledgeable employees. In a word, knowledgeable employees are groups who pursue independence, individualization, diversity, and innovation. Their motivation is more from the intrinsic reward of the work itself (Khan et al., 2021). However, there are generally two ways for enterprises to understand the satisfaction of knowledgeable employees with their salary. One is through a questionnaire, and the other is identifying employees’ emotions when they know their salary through models. The second method is widely used (Kurdi et al., 2020).

Based on the above background, this work first introduces the basic situation of the Convolutional Neural Network (CNN) and explains the basic working principles of CNN. Then, the advantages and disadvantages of the activation and loss functions in neural networks are described to facilitate...
the selection of different activation and loss functions according to different situations. Finally, in
the model, the attention mechanism is introduced to optimize this, and comparative experiments are
conducted to verify the rationality of this model and the relationship between employee compensation
satisfaction and job performance. The research innovation is to use CNN models to identify employees’
emotions in real-time, which is more accurate than the information obtained in the questionnaire
survey. This is because employees’ subjective feelings may occur during the questionnaire survey,
causing errors in the analysis results. Meanwhile, this work also provides direction for optimising
CNN models and contributes to the design of deep learning models.

Literature Review

Drosos et al. (2021) conducted interviews within enterprises to study employee salary satisfaction
and job performance. Based on the interview results and the company’s current situation of employee
satisfaction, they proposed research assumptions and built a research model. Then, questionnaire
distribution, data collection, data analysis, correlation analysis, regression analysis, and other methods
were adopted to verify the model proposed previously and identify the relevant influencing factors.
Finally, the factors significantly impacting employee satisfaction in the performance process were
identified, and influencing factors were informed to the company’s management for analysis (Drosos et
al., 2021). Liang et al. (2021) also researched through questionnaires, obtaining 410 valid questionnaires.
The qualitative comparative analysis method of fuzzy sets was applied to deeply explore the impact
mechanism of working conditions, sense of responsibility, leaders, and external rewards on employee
job satisfaction. The research results show that working conditions and leaders are vital in improving
employee job satisfaction, while responsibility and external rewards mainly play a “health care” role.
There are four practical ways to improve employee job satisfaction: work conditions and leaders are
the core conditions, and a sense of responsibility and external rewards are auxiliary conditions (Liang
et al., 2021). Regarding emotion recognition, Dzedzickis et al. (2020) proposed a subjective and
objective feature fusion neural network model for emotion recognition. This model could effectively
learn the spatiotemporal information of electroencephalogram (EEG) signals, dynamically integrate
EEG signals with eye movement signals, and output emotion classification results through a classifier.
Finally, comparative experiments were conducted using public datasets. Experimental results showed
that the new model’s accuracy was 86.27% and the standard deviation was 10.16%, which were superior
to traditional models. The new model could better utilize the complementary relationship between
subjective and objective features to achieve better emotion recognition effects (Dzedzickis et al., 2020).

In previous research, data were mainly obtained through questionnaire surveys to analyse the
relationship between employee salary satisfaction and job performance. This work studies the real-time
emotions of employees through neural network models. When optimizing emotion recognition models,
the previous focus was the information input, such as data preprocessing. However, this work modifies
the convolutional layer to improve the model’s performance and reduce the need for data processing.

ESTABLISHMENT AND OPTIMIZATION OF THE ANALYSIS
MODEL OF EMPLOYEE SALARY SATISFACTION

Convolutional Neural Network (CNN)

Deep learning is widely applied in multiple fields, especially in image classification. Thereby, CNN
is born. Its spatial invariance and channel specificity have significantly improved its image processing
speed and effect (Ghosh et al., 2020). CNN is a kind of artificial neural network which can divide the
image into several small areas during feature extraction and recognition. These small areas are called
“receptive fields”. The convolution kernel and receptive field size in CNN should be consistent. It
is necessary to multiply the pixel value of the image’s receptive field and the convolution bit by bit,
sum them, and then add the offset. Figure 1 displays the process.
The convolution operation calculates the input data by pushing the filter at a fixed distance. It is to multiply the filter data of each position and the corresponding input data, then sum them, and finally store the calculation results in a specific output location. The filter mentioned here is a convolution kernel, and the distance where the filter is applied is called the step (Ciancetta et al., 2020). Before the network model’s convolution operation, fixed data are sometimes added around the input data to adjust the output size. This operation is called a filling operation. It is a method often used in convolution operations, so the convolution operation can consciously adjust the output size and transmit the data to the next layer (Lou & Shi, 2020). Finally, the output data size can be calculated according to the input size, filling operation, and step size (Zhang et al., 2020). The calculation equation is as follows:

\[
H' = \frac{H + 2P - h}{s} + 1
\]

\[
W' = \frac{W + 2P - w}{s} + 1
\]

The size of input data is \((H, W)\), the filter size is \((h, w)\), the output size is \((H', W')\), the filling is \(P\), and the step size is \(s\). Most pooling operations refer to a down-sampling operation in the pooling layer, generally conducted independently on each channel. Its step size will refer to the shape and size of the receptive field used and normally plays a role in cooperation with the convolutional layer. It can make the network observe some features’ existence and reduce the space-range size and computation amount. Moreover, it can also prevent over-fitting and expand the receptive field (Chen et al., 2020). The most common operations in the network are maximum pooling and average pooling (Xu & Qiu, 2021). Maximum pooling refers to taking the maximum value in the selected range as the pooling result, and average pooling refers to taking the average value of all values in the selected field as the pooling result. Figure 2 shows the pooling operation.

The last is the fully connected layer. Its primary purpose in the deep network is to classify the network correctly. It is generally placed at the end of the model structure in practical use to weigh the features designed by various previous convolutional layers. It means integrating each component
and performing dimensionality reduction operations on the combined feature image (Khan et al., 2020). It does not specify the size of the input image, that is, as long as the image size is sufficient when the last layer of the network is input (Yang et al., 2021).

The confusion matrix is a kind of error matrix about image classification results. It is a commonly used auxiliary tool in deep learning. As its name implies, its function is to highlight whether there is confusion among multiple fundamental categories in the test set in the model (Markoulidakis et al., 2021). Generally, the matrix values involved in the confusion matrix are precision rate, recall rate, and accuracy rate. The precision rate is the probability that the real category accounts for the number of predicted column categories. It is the expected number divided by the total number of column categories (Qiu et al., 2020). Recall rate refers to the probability that the number of real values predicted by the model in the real value sample accounts for the real category. It is the expected number divided by the sum of row data (Shen et al., 2020).

The selection of the loss and activation functions in CNN is also significant. The smaller the value of the loss function is, the better the robustness of the network model is (Akbari et al., 2021). It is mainly applied to the training process of the network model. When the training data of each batch enter the network model, the output value is calculated using forward propagation. Then, the loss function will use a particular method to calculate the difference between the output value and the real value, which is the loss value (Clough et al., 2020). After obtaining the loss value, the model searches for the optimal weight parameter through back-propagation to achieve the purpose of model learning. It can reduce the loss between the real value and the predicted value so that the expected value generated by the model will approximate the direction of the real value. (Qu et al., 2022). The commonly used functions in classification tasks include the mean square error (MSE) and the cross-entropy error (Chen et al., 2021).

The MSE loss function is quite intuitive. It uses the Euclidean distance between the predicted and real values to find the loss value. When the predicted value is closer to the real value, the MSE of the two will be smaller (Gupta et al., 2020). The equation is as follows:

$$ E = \frac{1}{N} \sum_{k} (y_{(i)}^{\prime} - y_{(i)})^2 $$

$$ y_{(i)}^{\prime} $$ is the output of the neural network, $$ y_{(i)} $$ refers to the supervised data, $$ N $$ represents the sample quantity, $$ E $$ represents the loss function, and $$ k $$ represents the number of iterations. However, there is a disadvantage when using MSE functions. When the output probability value of its partial derivative
is close to 0 or 1, it may cause the partial derivative value of MSE to disappear when the network model is just trained. The partial derivative equation is as follows:

\[ E = \frac{2}{N} \sum_{i=1}^{N} (y_i' - y_i) y_i' (1 - y_i') x_{(i)} \]  

(4)

\( i \) is the dimension of the data, and \( x_{(i)} \) is the data input of the convolutional layer. The cross-entropy function is essentially a function derived from the application of cross-entropy knowledge in information theory in the field of communication to classification tasks. It is also a function often used in classification tasks. The smaller the value of cross-entropy is, the closer the two probability distributions are. Its equation is:

\[ E = -y \log(y') - (1 - y) \log(1 - y') \]  

(5)

\( y' \) represents the probability that the sample is predicted to be the true value, \( y \) refers to the sample's label, and \( \log \) is an exponential function.

In the network model construction, the activation function selection is also significant. It is a function that runs on the network neuron. Its function is to map the input to the output port in a nonlinear form, which increases the nonlinearity of the network model so that its network can be applied to various nonlinear models. It is the activation function (Wang et al., 2020).

The Sigmoid function is a composite function in exponential form, and it is the most frequently used function in the neural network. It tends to be close to human brain neurons in the physical sense and is the most familiar S-type function in biology. Its equation is:

\[ \text{Sigmoid} = \frac{1}{1 + e^{-x}} \]  

(6)

\[ \text{Sigmoid}' = \frac{e^{-x}}{(1 + e^{-x})^2} \]  

(7)

\( \text{Sigmoid} \) is the activation function, \( e \) is a constant, \( x \) is the data input, and \( \text{Sigmoid}' \) is the partial derivative of the activation function. Figure 3 displays its function image and partial derivative image.

When the input becomes increasingly larger, its output value gradually approaches 1, and it is smaller than 1 according to its output trend. As the input gets smaller and smaller, its output value gradually approaches 0, and the output value is always greater than or equal to 0 according to its output trend. In the back-propagation process, the gradient value will take the y-axis as the symmetry axis. Moreover, when the value of the Sigmoid function approaches 1 and 0, its gradient value tends to 0. That is, the gradient disappears (Szandala, 2021). The Sigmoid activation function is a composite function in exponential form. Its derivation involves the derivation of exponential function and the derivation of division. Hence, its calculation is relatively large compared with other functions, and the Relu activation function is an obvious piecewise function from the image. Figure 4 displays its function image and partial derivative image.

When the input is negative, the output is 0. When the input is positive, the output remains unchanged. Besides, when the Sigmoid activation function is back-propagated in the partial derivative diagram, its gradient value quite easily approaches 0. However, the Relu activation function can be regarded as a positive input value in the partial derivative process. The gradient is always 1, and there will be no gradient disappearance problem. The calculation equation is as follows:
Figure 3. Sigmoid function and partial derivative

Figure 4. Relu function and partial derivative
Relu = \max(0, x) \quad (8)
Relu' = 1, x > 0 \quad (9)
Relu' = 0, x \leq 0 \quad (10)\]

Relu is the activation function, \( x \) is the input data, and Relu' is the partial derivative of the activation function. After the basic situation of CNN is discussed, the next step is to optimize the emotion recognition model to improve the recognition accuracy.

**Facial Emotion Recognition Algorithm Based on Attention Mechanism Network**

Attention is generally an idea that tends to focus on aspects or certain areas. When attention is added to the neural network, the network will become quite intelligent. When it extracts features, it will focus on extracting features of key parts. However, ordinary CNN aims only at global features and cannot selectively pay attention to salient features of key positions for input images (Cai & Wei, 2020). Compared with the general convolution’s facial expression recognition algorithm network, it is based on capturing the overall global features of the face. With the proposed facial expression algorithm based on the attention mechanism, it is easier to capture the key information of the face part, namely the local information of the facial expression. It combines attention to focus the network on key parts of the face (Qi et al., 2020). Face emotion recognition based on the attention mechanism mainly includes three modules: preprocessing, feature extraction, and expression classification (Jang et al., 2020). Figure 5 displays the specific process.

This work improves the attention neural network, which mainly consists of three parts: space converter, backbone network, and attention network. Figure 6 presents the optimization results.

The space converter network makes the network model have space invariance. It corrects the direction of the image by letting the neural network learn the spatial changes of the image, such as cutting and zooming a side face. It turns the image into the ideal direction of the face, performs a spatial transformation on the input image, and outputs a new image (Wen et al., 2022). The space converter network is mainly composed of two parts. One part is the local network, and the other part is the grid generator (Xu et al., 2022). This work designs a local network of two fully connected layers, as shown in Figure 7.

---

Figure 5. Facial expression recognition
The dataset involved here is small, and the requirements for network complexity are low. Moreover, to highlight its effect, the backbone network designed in this algorithm is improved based on the Alexnet model (Li et al., 2020). Generally, the Alexnet model comprises five convolutional layers and three fully connected layers with eight layers of network structure. Overlapping pooling is selected as its pooling method. The size of the pooling window of this pooling method is larger than the step size, so each pooling has overlapping parts, which can avoid partial over-fitting (Hao et al., 2022). This work optimizes the network by adding the asymmetric convolutional layer. Figure 8 displays the optimized network.

The replacement by the asymmetric convolutional layer enhances the backbone and does not require additional parameters. Channel attention is used to model the importance of each channel and then highlight or constrain channels for different tasks (Kong et al., 2021). The specific process is as follows. The input feature first passes through the global max pooling layer and the average pooling layer. Then it passes through two 1*1 convolutional layers to increase and reduce the dimension.
Finally, two vectors with the same dimensions are generated and added to generate the channel through
the activation function (Shi et al., 2022). Figure 9 is its channel attention model.

Spatial attention is mainly used to improve the feature representation of key parts. The specific
process is as follows. Two feature maps are obtained through the average pooling and global max
pooling of channels for input features and then are spliced based on channels. The feature map after
the cascade undergoes a 7*7 convolution operation. Its purpose is to reduce the dimension of the
feature map, generate spatial attention through the activation function, and finally obtain the scaled
new features (Wen et al., 2020).

The attention module here adopts global average pooling and global max pooling, which is biased
towards the overall situation. However, the attention mechanism network used is based on the above
attention mechanism. It uses the importance pooling method to form local channel attention and
cascades with the above spatial attention to finally form the improved attention mechanism module. The
importance pooling method is a pooling layer based on local importance, highlighting the importance
of different features. Compared with global pooling, it can highlight classification features more.
This work applies it to channel attention to form channel attention of local importance. The process
of adding importance pooling to channel attention is quite simple. First, the input features generate
the importance map through convolutional learning. Then, its importance is applied to perform the
normalization operation. After that, the obtained eigenvalues are multiplied by the input eigenvalues
after exponential calculation. Finally, the division operation with the eigenvalues after the exponential
operation should be conducted to better learn the importance and get channel attention. Figure 10
displays the improved attention module.
Figure 9. Channel attention mechanism

Figure 10. Improved attention module
When training the model, the data first pass through the space converter network, that is, after affine transformation; the input eigenvector is still 1*48*48. Finally, the attention module is added to the first and last convolutions of the model.

**Design of Questionnaire for Knowledgeable Employee**

Knowledgeable workers are professionals who focus on knowledge and are driven by innovation. They possess a high degree of autonomy and creativity and can solve problems and promote organizational development through the use of knowledge. This type of employee usually has a high degree of education and professional skills, and is mainly engaged in knowledge-intensive or highly-informative fields, such as research and development, design, consulting, and marketing. Unlike workers in the traditional sense, knowledgeable employees pay more attention to exerting their creativity and value and focus more on self-realization and growth. Therefore, enterprises need to provide these employees with a good work environment and promotion mechanism and stimulate their enthusiasm and creativity to ensure their continuous contribution to the organization and competitive advantage. Hence, understanding their satisfaction with salary becomes quite crucial. Based on this, this work designs a questionnaire, and the answers to the fill-in-the-blank questions will be artificially classified. Table 1 is part of the questionnaire.

**EXPERIMENTAL RESULTS AND ANALYSIS**

**Comparative Experimental Results of the Emotion Recognition Accuracy Between the Optimized Model and the Traditional Model**

Seven emotions are divided into three categories: positive, negative, and neutral. The selected dataset is the public FER2013 dataset. Table 2 shows the equipment environment used in this experiment.

Figure 11 is the experimental result of the recognition accuracy comparison between the traditional and the optimized models. Figure 11a shows the data from the conventional model, and Figure 11b presents the data from the optimized model.

Figure 11 suggests that the traditional emotion recognition model has the highest recognition rate for neutral emotions, and its recognition accuracy is 81%. Its accuracy of positive and negative emotion recognition is only 76% and 77%, respectively. The optimized model’s recognition rate of

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Concrete problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction with salary level</td>
<td>Your actual income</td>
</tr>
<tr>
<td></td>
<td>Your monthly salary</td>
</tr>
<tr>
<td></td>
<td>What is your monthly salary level in the company?</td>
</tr>
<tr>
<td></td>
<td>What is the monthly salary level for your current position in the entire company?</td>
</tr>
</tbody>
</table>

**Table 1. Salary satisfaction questionnaire for knowledgeable employees**

<table>
<thead>
<tr>
<th>The server</th>
<th>Dell EMC R740</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running memory</td>
<td>128G</td>
</tr>
<tr>
<td>Disk</td>
<td>16T</td>
</tr>
<tr>
<td>Operating system</td>
<td>Centos7.6</td>
</tr>
<tr>
<td>Computer voice</td>
<td>Python</td>
</tr>
</tbody>
</table>

**Table 2. Equipment environment**
neutral emotion is 95%, the highest. Moreover, its recognition accuracy for positive and negative emotions is 92% and 91%. It reveals that the accuracy of neutral emotion recognition of the two models can differ by up to 14%. This comparative test verifies the effectiveness and feasibility of the optimized model.
An Experimental Analysis of the Effect of Salary Satisfaction of Knowledgeable Employees on Job Performance

The emotion of employees when they know their salary can reflect their satisfaction with the salary. The salary data and job performance data of 200 actual employees are selected. Besides, the company has done a salary satisfaction questionnaire. When employees know their salary, the proportion of employees with positive, neutral, and negative emotions is 73%, 21%, and 6%, respectively. Job performance is the January performance. Performance is divided into two levels. Performance ranking in the top 50% is excellent, and performance ranking in the bottom 50% is good. Figure 12 presents the experimental results.

Figure 12 shows that 93 employees who maintain positive emotions have excellent performance, accounting for 63.7% of the total number of employees with positive emotions. Among the employees whose performance ranks in the bottom 50%, 9 people are in a negative mood, accounting for 75% of the total number of people in a negative mood. It suggests that regarding salary satisfaction, the performance of employees with positive emotions is generally relatively high, while that of employees with neutral emotions is mixed. However, the performance of employees in a negative mood is mostly in the bottom 50%. It reveals that the salary satisfaction of knowledgeable employees will impact their job performance.

DISCUSSION

The research results show that the optimization model is significantly higher than the traditional model regarding recognition accuracy. The recognition accuracy rates for the three emotions are 92%, 95%, and 91%, with all reaching over 90%. However, the recognition accuracy of traditional models is 76%, 81%, and 77%, respectively, which is much lower than the optimization model. The main reason is that...
the optimization model introduces an attention mechanism, making it more accurate in capturing key facial information. Meanwhile, a new fully connected layer is added to the optimization model, which processes the collected information multiple times, and then the optimization model can collect new feature information. The analysis of the relationship between salary satisfaction and job performance of knowledgeable employees reveals that the higher the salary satisfaction of employees is, the higher their job performance is. There is a positive correlation between the two. This is because most job performance for knowledgeable employees comes from completing corresponding prescribed work tasks on time. The higher their salary satisfaction is, the more motivation they have. The emotion recognition research conducted by Li et al. (2022) improved recognition accuracy through data preprocessing, with the highest recognition accuracy being only 88.3%, which was lower than the average recognition accuracy here. Moreover, the research conducted by Islam et al. (2021) optimized recognition accuracy by adding an attention mechanism. However, compared with this model, the model proposed here adds fully connected layer optimization. Although the recognition speed of the proposed model is slightly slower, its recognition accuracy is higher than the model being referenced (Islam et al., 2021).

CONCLUSION

With the times’ progress, enterprises’ demand for knowledge-based talent has become increasingly higher. How to retain employees and make them play their best role and use their best ability in their jobs have become issues that enterprises need to consider. Therefore, exploring the relationship between employee salary satisfaction and job performance is crucial. This work first introduces the details of the CNN model, loss function, and activation function. Then, the CNN model is optimized by introducing an attention mechanism to improve the recognition rate of the emotion recognition model. Finally, through comparative experiments, the rationality of the model is verified and the impact of employee salary satisfaction on employee job performance is verified. The experimental results show that the recognition accuracy of the model proposed is higher than that of the traditional model, especially the recognition rate of neutral emotions, which can reach 95%, thus verifying the effectiveness and feasibility of the model proposed. Additionally, the optimized model is adopted to carry out an experimental study on the salary satisfaction and job performance of employees in a company. It is found that 63.7% of the employees who are in a positive mood and have high satisfaction with their salary have excellent performance. Employees in a negative mood have low satisfaction with salary, and 75% of them have good performance, indicating that their satisfaction with their salary will directly affect their job performance. However, this work also has some deficiencies. First, in addition to salary satisfaction, multiple other factors also affect employee performance. Other variables will be added in the follow-up study. Second, although the optimized model improves the recognition accuracy, it makes the model operation more complex and increases the amount of calculation. Later, the calculation speed of the model will be improved by reducing the number of modules.
REFERENCES


APPENDIX

Salary Satisfaction Questionnaire

Dear Sir/Madam,

Hello! Thank you very much for taking the time to fill out this questionnaire. This questionnaire is purely academic and mainly aimed at understanding your current salary satisfaction. It is answered anonymously, and the answers you fill in will not be leaked to any individuals or companies other than yourself. Your answers are used only for academic research analysis. Please answer the questions below based on your actual situation.

Thank you for your cooperation, and I wish you all the best!

Please fill in the options that meet your criteria in brackets for questions 1-5.

1. Date of birth: ( )
2. Gender: male ( ) female ( )
3. Education level: junior college () undergraduate () graduate or above ()
4. Your working experience: less than one year () 1-2 years () 2-5 years () 5 years or more ()

Questions 6-18 are fill-in-the-blank questions. You can fill these in according to your actual situation.

5. Your actual income:
6. Your monthly salary:
7. Your monthly salary level in the company:
8. The current monthly salary level for your position in the company:
9. All of the benefits you enjoy:
10. The benefits and remuneration paid by the enterprise to you:
11. The value of the benefits you enjoy:
12. The number of benefits you receive:
13. Your opinion on the recent salary increase:
14. The most representative salary increase you have received in the past:
15. Factors affecting salary increases:
16. The consistency of the company’s compensation policy and implementation:
17. Salary difference of employees in different positions of the enterprise:
18. Enterprise salary management mode:

Please judge based on your actual situation for questions 19-24. If you meet the criteria, please write 1 in the brackets following the title.

19. Can you complete your tasks on time? ()
20. Can you effectively achieve your work objectives? ()
21. Is teamwork possible at your place of employment? ()
22. Are you very efficient? ()
23. Do you have a good relationship with the supervisor? ()
24. Are you willing to share the work of others? ()

This is the end of the questionnaire. Please check whether there are any missing items. Thank you again for your support!