Journal of Electronic Research and Application

Research Article

Voice Interaction System for Race Games

Xuanning Liang

North Raleigh Christian Academy, North Carolina, United States.

Publication date: December, 2019

Publication online: 31 December, 2019

****Corresponding author:*** Xuanning Liang, liyucheng@xyzrgroup.com

1 Introduction

With the advancement of the computer technology, there are now various kinds of computer games, especially race games (RAC). However, there are still many issues in race games. The first problem is that the interactivity is not fully-fledged. Information can only be input to the computers by manipulators without any feedback from the computers. The second problem is that players cannot pay attention to the information about the road conditions while they are driving. The main reason is that the monitor is usually too small to show the 3-D(three-dimensional) scenes of the game and it is hard for players to notice the road conditions. In order to figure out these problems, we designed a voice interaction system to improve the game experience.

In the future, this technique will have a bright prospect. On one hand, this invention can be applied to VR games which will be one of the most popular research directions. While on the other hand, this invention can also be applied to automatic drive which will definitely change people's daily life.

2 Procedure

Step1: Encoder

For the computers are unable to recognize the object in the picture like human, instead, computer recognizes the picture as two-dimension matrix, the difficulties we faced during the research project at first is how to let the computer understand different object in the picture. To solve the problem, we use the Convolutional Layers from Faster R-CNN, which is advanced convolutional neural network in deep learning, to help computer acquire the feature of the computer. Then, in the activate Layer, the feature will be determined the anchor whether it belongs to negative or positive through Softmax algorithm in PRN (Region Proposal Networks), and the bounding box regression will revise the anchors in order to achieve the precise proposal. Afterwards, Pooling Layer will collect the feature map from convolutional Layer and proposal from PRN to compound proposal feature maps. Finally, it will classify the proposal through proposal feature maps and obtain the accurate location.

WMD (Word Mover's Distance) can model the distance between two documents as a combination of semantic distances for words in two documents. The computer trains this process for specific time, and will realize the similarity between two words. So, if you input something, the computer will output the most similar answer if there is no answer matches the players' question.

WMD formula: $\sum_{i,j=1}^{n} T_{ij} c(i,j)$

Through Word2vector and WMD (Word Mover's Distance), the question and history can be understood by the computer. For each sentence, we only allow twenty words at most. We will add "0" at the end of sentence or delete the extra word if the sentence is not exactly 20 words in one sentence. Then, this sentence will go through the LSTM (Long Short-Term Memory) process. During this process, the computer will filter the useless information, remember the essential information and decide what will output.

Step2: Decoder

Firstly, this group changes two datasets into the same size in order to combine them. We collect the Encoder input data, which size is (32×10×512). The batch

size is 32, we change the size of (32×10×512) into (32×10×100×512). It means that we test 32 image one time, each image will be asked 10 round and each round has 100 answers. The max length of each answer is 512. Then, we change the input from (32×10×100×512) to (32000×512) by:

Batch size×rounds×number of answers= 32×10×100=32000

After that, we should preprocess the answer into vector and change its size. In the beginning, we have to embed the answer by semantic Layer. The internal structure of semantic Layer can be seen in Figure 4. We set the embedding size as 300. We input the answer into ELMo and Glove model respectively, the output vector from ELMo would go through a Fully Connected Layer to change the size from 1024 to 300. We contact these two 300 dimensions vector into a 600 dimensions word vector. The size of this vector is (32×10×100×20×600) Next, we establish the LSTM Layer3, we put the vector from semantic Layer into this Layer. The size will change from (32000×20×600) to (32000×512).

Secondly, we combine these two data in Answer score module.

Thirdly, we use Softmax activation function to get the weights of all the 100 answers. The Softmax function is shown as below:

$$\sigma_i(z) = \frac{e^{Z_i}}{\sum_{j=1}^{m} e^{Z_j}}$$

Since we use the supervised learning, we know the right answer, we set the right answer as 1, wrong answers were set as 0.

Fourthly, we use Cross Entropy Loss Function to calculate the loss. Cross Entropy Loss Function describes the distance between two probability distributions, and the smaller the cross entropy is, the closer they are to each other. The formula is shown as below:

$$\mathrm{C} = -\frac{1}{n}\sum_{x}[y\,lna + (1-y)\ln(1-a)\%$$

We set 'a' as weight for each answer and 'y' to be 0 or 1 (0 for wrong answer or 1 for right answer). We can obtain the loss value by using it.

Finally, pass the loss back to change the parameter and retrain the model.

3 Testing results

Learn from the work of our predecessors, we use six parameters R@1,R@5, R@10, Mean, Mrr and Ndcg as the indexes to evaluate this model.

From 10 epochs evaluation data, we find the recall rate, Mrr and Ndcg is increasing, the value of mean is decreasing, which indicates that with the training of the data, the accuracy of the algorithm is improving. As a result, this whole model is valid.

4 Conclusion

In order to work out the problem that players cannot take the proper action during a race game due to the restriction of the monitor, our design proposes a voice interaction system for race games based on deep learning. The design can complete the extraction of the image information and realize the interaction with the player, thereby improving the game experience of the player.