Datathon Documentation

Team TechCrew

Approach Used

We have developed a neural network for classification in order to recommend places to users according to their preferences and positive attributes of the destinations and our model predicts a diverse set of destinations that align with user's preferences well.

Data Preparation and Preprocessing

1. **Data Cleaning**

   Included preprocessing steps to handle missing values, outliers, and inconsistencies in the data to ensure high-quality input for the model.

2. **Feature Engineering**

   Involved creating new features and modifying existing ones to enhance model performance, including the extraction of relevant characteristics from raw data.

3. **Activity Destination Weights**

   Determined weights for various activities and destinations to improve recommendation accuracy by emphasizing user preferences and destination attributes.

4. **User Ratings and Reviews**

   Analyzed user ratings and reviews to classify sentiment and calculate destination scores, which were then used to refine recommendations.

5. **Standard Scaler**

Used in order to scale features for a better performance and convergence speed of the model due to the same scale.

Model Creation

We used a neural network for this recommendation system due to its benefits like non-linear relationships, scalability, high-dimensional data handling ability and the ability to learn from sparse data.

Dense layers were used to build the network and the BatchNormalization was used to normalize activation in order to improve training speed and stability and the dropout layers to prevent overfitting.

The activation functions used were ReLU for hidden layers and the sigmoid function for the output layer and the Adam optimizer was also used to adjust weights during training. Binary cross-entropy was used due to multi-label classification as the loss function.

Training and Validation

1. EarlyStopping

   Used to stop the training when the validation loss stops improving to prevent overfitting.

2. ReduceLROnPlateau

   Used to reduce learning rate when a plateau in validation loss is detected since it adjusts the learning rate dynamically to help the model converge more effectively.

Model Evaluation Metrics

1. Classification Report

   1.1 Precision

   - Reason - Precision measures the proportion of correctly recommended places out of all the places recommended by the model. So high precision indicates that the recommended places are likely to be relevant and align well with the user's preferences. This ensures that users receive recommendations that are specifically suited to their interests.

- Relevance - By maximizing precision, we have ensured that the recommended places are indeed the ones that are most relevant to the user's preferences.

1.2 Recall

- Reason - Recall measures the proportion of relevant places that are correctly recommended out of all relevant places available. High recall indicates that the model is effectively identifying most of the relevant places that match the user's preferences. This means the model is good at capturing the full range of preferences the user has.

- Relevance - Maximizing recall helps ensure that the model does not miss out on relevant places, providing a more comprehensive list of recommendations. Higher recall can indirectly support diversity by ensuring a broad set of relevant places are considered, increasing the chances of diverse recommendations if relevant places span various categories.

1.2 F1 Score

- Reason - The F1 score is the harmonic mean of precision and recall, providing a single metric that balances the two. The F1 score helps to gauge the overall performance of the recommendation model, taking into account both how accurate and how comprehensive the recommendations are.

- Relevance - By balancing precision and recall, the F1 score ensures that the recommendations are both relevant and comprehensive. It prevents the model from being overly biased towards precision (too conservative) or recall (too broad). The F1 score indirectly contributes to diversity by encouraging a balance between recommending highly relevant places and covering a wide range of relevant options.

2. Weighted F1 Score

- Reason - The weighted F1 score averages the F1 scores across different classes (places) and weights them according to the number of true instances for each class. This metric is particularly useful in scenarios where some classes (places) are more common than others. It provides a balanced measure of performance across various classes, considering their frequency.

- Relevance - The weighted F1 score helps ensure that the model performs well across both frequent and infrequent places, making sure that all relevant places, regardless of their frequency, are considered appropriately. By considering the frequency of different places, the weighted F1 score ensures that both popular and less common places are included in the recommendations, promoting a diverse set of suggestions.

Refinement for Destination Diversity

Ensured that the model includes at least one destination from each activity category to promote a diverse set of recommendations.