Mit dieser Folie möchte ich die bisherigen Ergebnisse darstellen.

Bis letzten Treffen haben wir schon viele neuen Attributen von Polygon extrahiert und ein Seq2seq Model für die Vorhersagen der Bewertung gebaut. Und mit ??% accuracy in Top5 können wir sagen, dass die Bewertung ist vorhersagbar.

Aber es gibt noch ein Problem bei diesem Model. Die Anzahl von Polygon ist viel zu weniger als Real. Ob das Model benutzbar ist oder wie wir das Model ändern können, wenn wir viel mehr Polygone haben, ist die Hauptaufgabe, die wir in letzter Woche bestätigen muss.

Um die Aufgabe zu lösen, designieren wir ein Part-Embedding Phase. Also representieren wir jedes Polygon mit einem Vektor. Hier wir benutzen Word2Vec Model. Gebe wir das Polygon ein, bekommen wir als Ausgabe einen Vektor. Das Ziel hier ist, für die ähnlichen Polygone, ähnliche Vektor haben sie.

Um es zu schaffen, sollen wir selbst ‚Sätze‘ erfinden. Und mit diesen ‚Sätze‘ trainieren wir das Word2vec Model.

Anhand der Attributen wir bekommen, machen wir zuerst Clustering, und in jedem Cluster wähle wir Polygon zufällig aus. Aus diesen Polygone bilden wir ‚Sätze‘.

Haben wir die Seq2seq Model nicht geändert. Dieses Bild zeigt die Funktion vom seq2seq Model. Als Eingabe haben wir 30 Platte, jede Platte enthaltet 50 Polygone mit der bestimmten Reihe. Seq2seq bewertet dann jede Polygon auf den Platte.

Hier zeigt die grundsätzliche Information für das Experiment. Wir benutzt 70,000 Daten für die Trainieren des seq2seq Models und 8,000 Daten für Testen. Die Struktur des Models ist einfach ein LSTM Network.

Für das Testen des Models haben wir zwei Metriken designiert. Die erste Metrik haben wir bei letzten Treffen schon erklärt. Eine Seite berechnen wir ‚Mean score‘ für jede Platte und sortieren wir diese Platten nach ‚Mean score‘, holen wir dann Top 5 ab. Andere Seite wähle wir die Platte aus, die die wirkliche minimale ‚Mean score‘ hat, hier wir als ‚mean target‘ genannt. Und schau mal ob die ausgewählt Platte in Top 5, wenn ja machen wir Hit plus eins, sonst machen wir Miss plus eins. Hit durch (Hit + Miss) ist der Score des Models, Hier bekommen wir accuracy mit 82%

Im zweiten Metrik, nachdem Sortierverfahren, behalten wir die Reihenfolge der Platten. Andere Seite wähle wir die besten Platte aus, wie im letzten Metrik. Und notieren wie die Position der ausgewählten Platten in die sortierte Reihenfolge der Platten.

Dieses Bild zeigt die Ergebnisse vom Metrik2. Es heißt von 80 Versuchen, 35 sind in Position 1. 17 sind in Position 2 und 7 in Position 3.