

Sometime the flexible transformationn are emergency

A cute kid.

Abstract

In the real world, different expressions of human language often convey different intentions. Existing intent recognition datasets typically provide a given text segment and then let the machine identify the intent corresponding to that text. This paper introduces a baseline method for an intent recognition dataset to address this issue. It first builds the corpus based on information gathered from both in-scope and out-of-scope data. For the different samples, I employ an attention mechanism to capture the overall semantic meaning. Then, I utilize two matrices to identify the intent within the semantics. The dataset contains 23,700 high-quality samples, covering 151 intent classification labels across 11 domains. I trained the model on this dataset and conducted a series of analyses, including PCA, attention visualization, probability distribution analysis, and discrepancy analysis. Furthermore, I experimentally analyzed the differences in model performance obtained from training datasets of varying magnitudes. My research findings indicate that intent recognition in the field of machine learning differs from intent recognition in the real world. And I hope this baseline method will advance related research efforts.

Keywords: *Intent recognition, machine learning, attention mechanism*

I. Introduction

Intent recognition is crucial in natural language processing, which aims to leverage textual information to determine intent categories for improved communication interactions. Although existing text-based intent recognition primarily focuses on goal-oriented tasks within specific domains, these tasks typically originate from instructions or queries possessing distinct semantic features. After a true understanding of such natural language processing tasks will enhance human performance in the real world and advance the development of tasks in other domains, for example, the multimodal language that incorporates richer information about emotions, attitudes, and behaviors. Figure 1 demonstrates the process of intent recognition. Given a segment

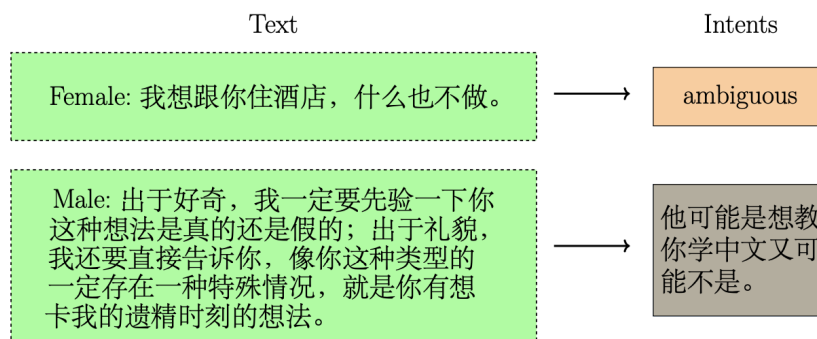


Fig. 1. Introduction.

of text, let the model recognize the intent of this text. In this paper, I propose a base line method for this task. The main contributions are followed:

- The performance of this model was analyzed from the perspective of data volume.
- Perform dimensionality reduction analysis on the model's output.
- Use violin plots to analyze the probability distributions of the model.
- Visualize the model's attention outputs using heatmaps.
- Use scatter plots to analyze the differences in the model's output.

II. Dataset

I use Clinc150 dataset. This dataset includes 150 domain-specific intent labels and one out-of-scope label. This dataset contains a total of 23,700 records. I designate the records marked with “train” to the training data, with the remaining records forming the test set.

III. Model

Figure 2 demonstrates the architecture of the model. Given a batch of data, the model uses its pre-trained vectors to embed the data which is Glove[1]. Then, the model feeds the embedded representations into the attention mechanism. The attention mechanism extracts the first token from each data point and performs a matrix multiplication with the overall representation of the current data to obtain the corresponding attention weights. The obtained attention weights first undergo a softmax operation, followed by another matrix multiplication with the overall representation of the current data, thereby generating the final attention output. Then, the attention output is passed to the first matrix for matrix multiplication. The resulting output is

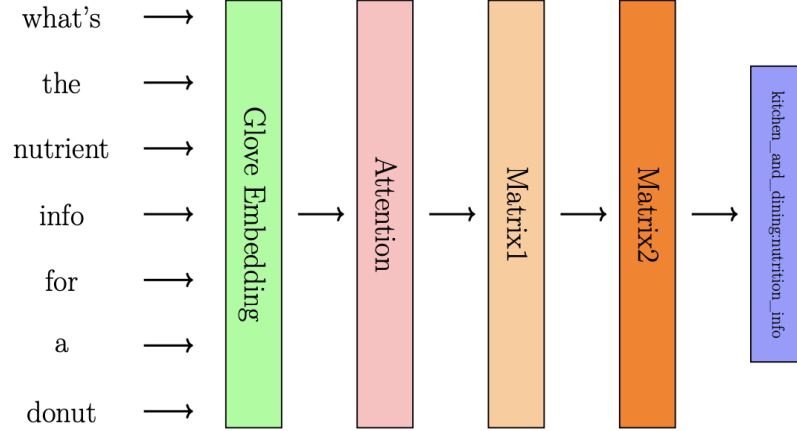


Fig. 2. The atchitecture of the model.

then fed into the second matrix for correspondding label classification.

IV. Results

Table I demonstrates the hyperparameters of the model. Table II shows the performance of models trained with different amounts of data. Figure 3 illustrates the relation of training epochs and loss value across different models, along with the corresponding line charts in Table II. It can be observed that the convergence behavior of different models is independent of the amount of data. Furthermore, with the volume of training data increases, model performance improves; however, insufficient data leads to a decline in model performance.

TABLE I. hyperparameters.

hyperparameters	value
torch.manual_seed	66
Adam learing rate	0.001

TABLE II. Main Results.

Data volume	Accuracy
0	$\frac{75}{8700}$
10	$\frac{51}{8700}$
100	$\frac{475}{8700}$
1000	$\frac{2042}{8700}$
10000	$\frac{4334}{8700}$
15000	$\frac{5214}{8700}$

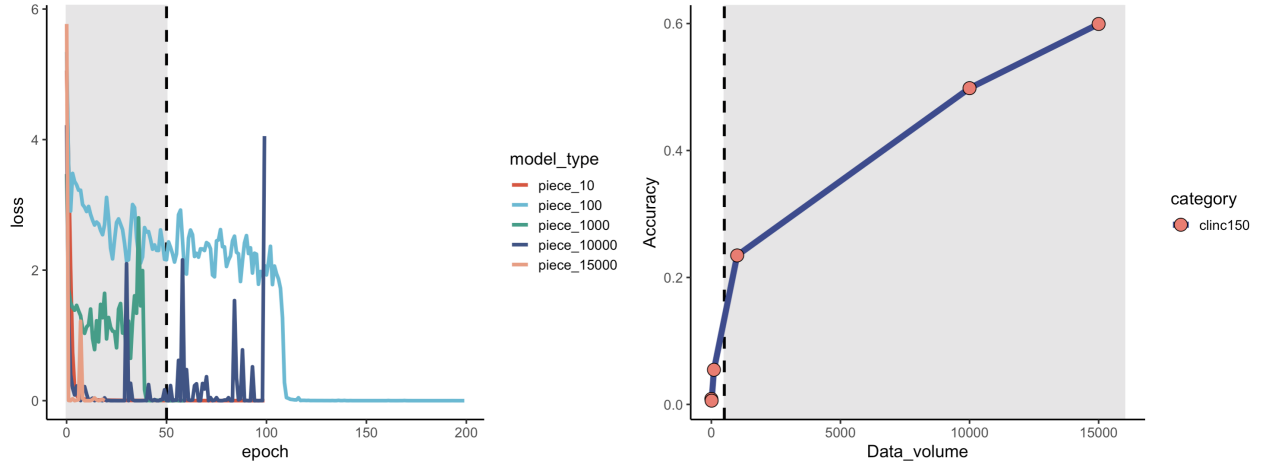


Fig. 3. Models accuracy and their training loss.

V. Principal Component Analysis

Figure 4 shows the model output after dimensionality reduction. Obviously, within the volume of data increases, the model's output becomes increasingly organized and gradually takes on a triangular shape.

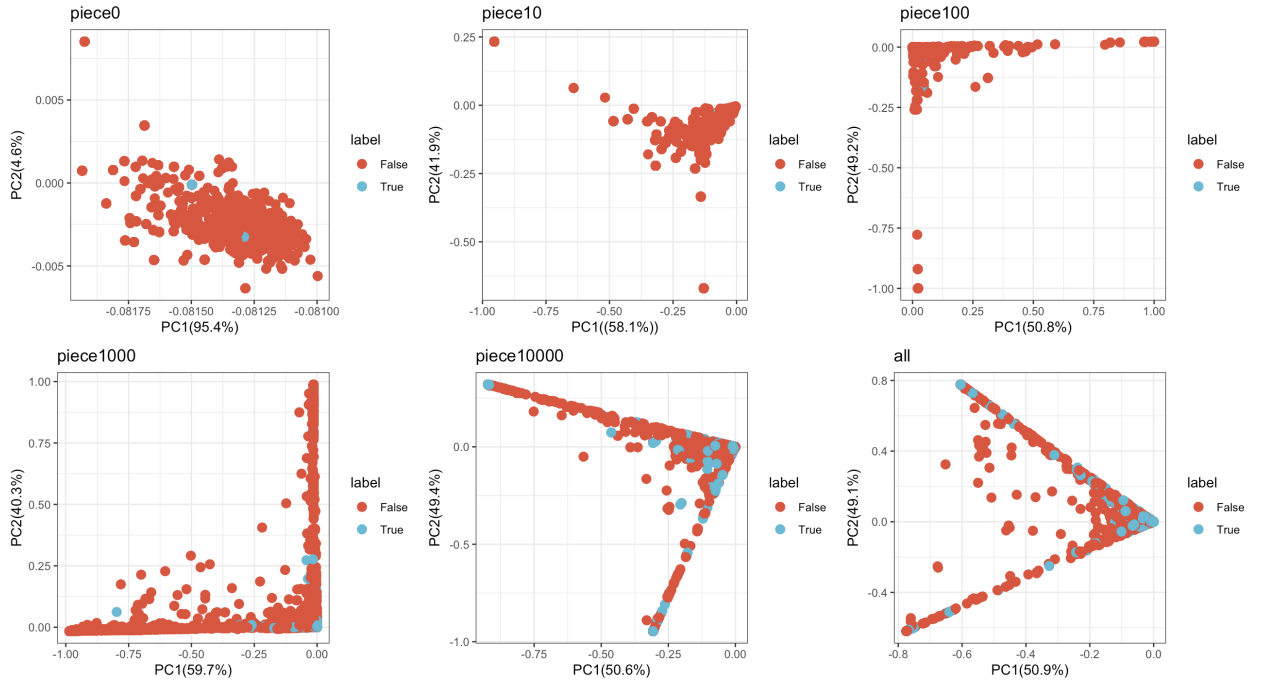


Fig. 4. PCA.

VI. Probability Distribution Analysis

Figure 5 displays a violin plot of the logical values corresponding to the model's true labels. The pink portion is constructed from logical values via a density function. The three lines in the box plot, from top to bottom, represent the upper quartile, the median, and the lower quartile. Points within the distribution form an enhanced jitter scatter plot, where the width of the jitter is controlled by the density distribution of the data.

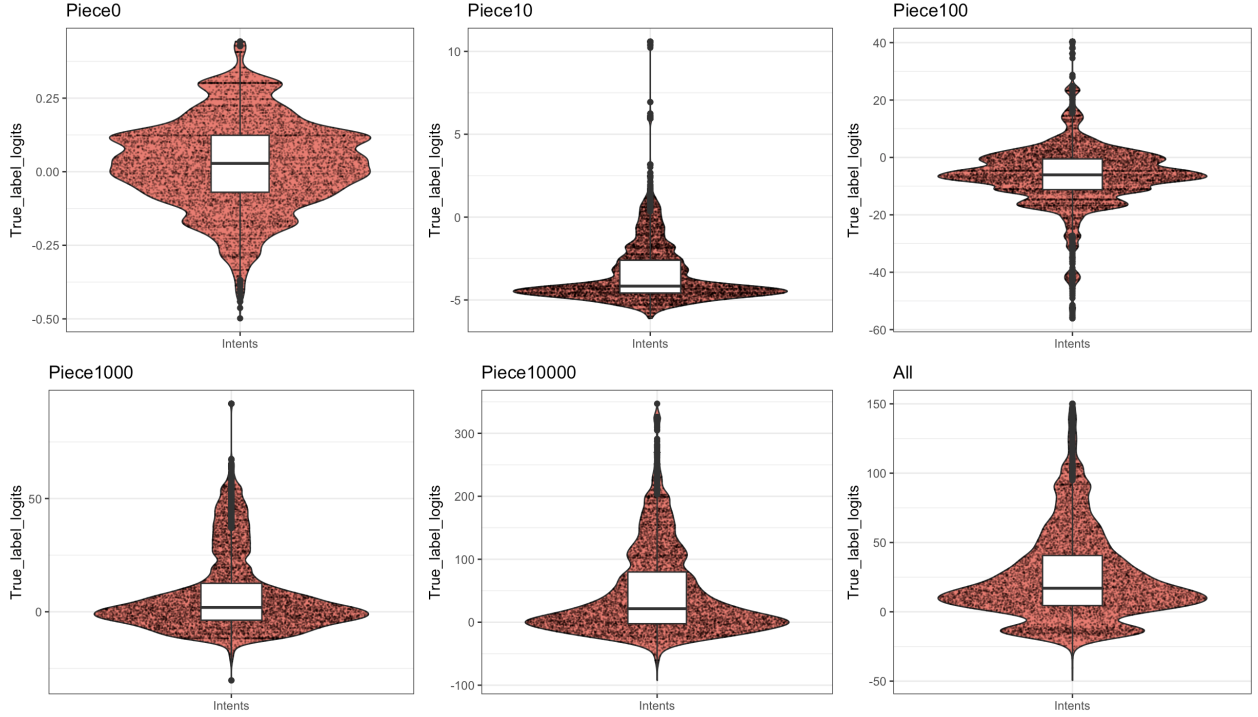


Fig. 5. Probability Distribution.

VII. Attention Visualization

Figure 6 shows a heatmap of the model's attention weights. From the perspective of color recognition, despite having different parameters, models with less training data exhibit higher similarity in their parameters.

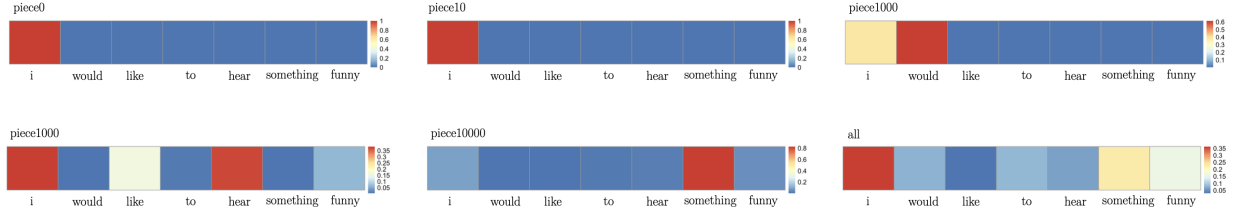


Fig. 6. Attention visualization.

VIII. Discrepancy analysis

Figure 7 shows a scatter plot for the discrepancy analysis. The x-coordinate values of the points in the figure follow equation 1. More details, values greater than zero shift two to the right, while values less than zero shift two to the left. The RD_value

$$\log_2 \text{foldChange} = \log_2 \left| \frac{\text{pred_logits}}{\text{true_label_logits}} \right| \quad (1)$$

follow equation 2. The 'RD' is the short of restrain desire. The meaning expressed by the corresponding formula is to allow the points in the diagram to diverge more widely while simultaneously forming a regular shape.

$$\text{RD_value} = 20 * \sin\left(\frac{\sum_{i=0}^{\text{total_classes}} w_i}{\sum_{i=0}^{\text{total_classes}} |w_i|}\right) \quad (2)$$

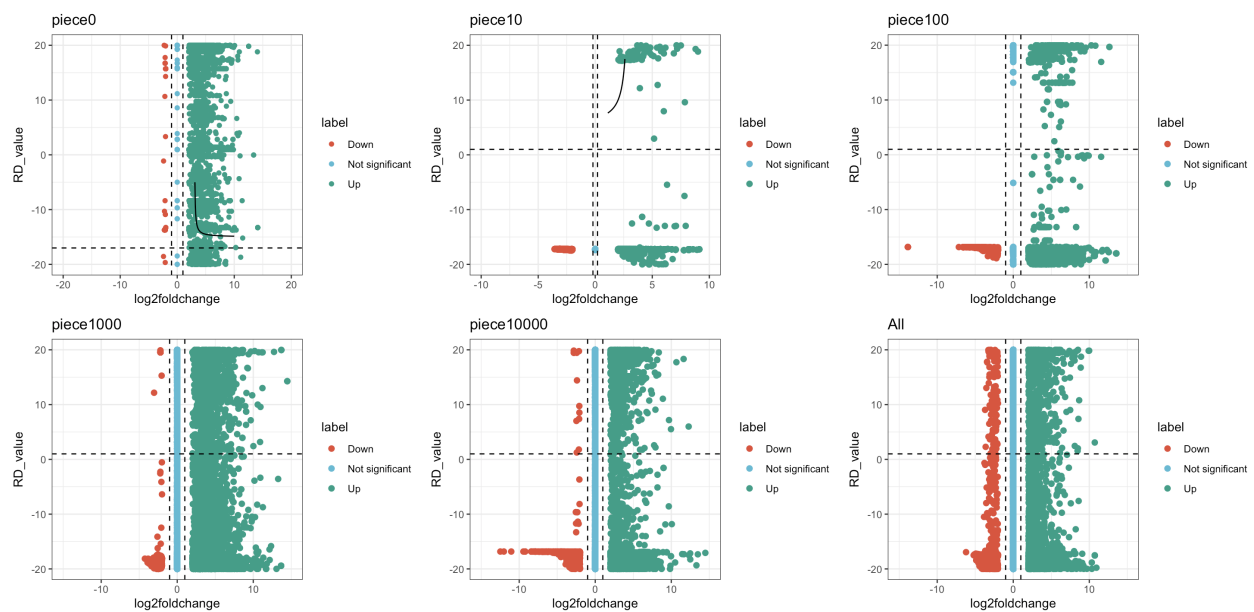


Fig. 7. discrepancy analysis.

IX. Conclusion

This paper provides a brief analysis of the baseline method for intent recognition from the perspective of training data volume.

References

- [1] *Glove: Global vectors for word representation*, Pennington, Jeffrey and Socher, Richard and Manning, Christopher D, Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP).