【word2Vec.py】

该代码利用gensim库，用五个临床医疗句子作为训练语料来训练Word2Vec模型，将词语映射为向量。演示了如何计算词语相似度等分析，还借助scikit-learn进行PCA降维，并用matplotlib将词向量在二维空间中的关系可视化为散点图。

Using the gensim library, this code trains a Word2Vec model on a corpus of five clinical medical sentences to map words into vectors. It not only demonstrates analyses like calculating word similarity but also leverages scikit-learn for PCA dimensionality reduction and matplotlib to visualize the word vector relationships as a 2D scatter plot.

【ICU\_prediction.py】

该代码应用LSTM模型对模拟的ICU血压的单变量时间序列进行预测,然后利用matplotlib将原本的血压数据、模型在训练集和测试集上的预测结果一起绘制出来。

Using an LSTM model, this code forecasts a simulated univariate ICU blood pressure time series and then visualizes the results with matplotlib, plotting the original data against the model's predictions for both the training and test sets.

【biobert\_classifier.py】

该代码利用BioBERT嵌入向量完成医学文本分类任务。首先使用BioBERT模型，将Excel文件中的医疗病历文本转化为能被机器理解的数字特征向量。随后，基于这些向量训练逻辑回归分类器，用以自动判断病历应该分到哪个科室。最后，通过PCA降维和matplotlib绘图，将文本向量在二维空间中可视化，直观展示不同科室文本在特征空间中的分布情况。

This code leverages BioBERT embeddings to classify medical texts. It begins by converting medical notes from an Excel file into numerical vectors using the BioBERT model. A logistic regression classifier is then trained on these vectors to automatically sort the notes by department. Lastly, it employs PCA and matplotlib to create a 2D plot, visualizing how the texts from different departments are clustered in the feature space.

【note.xlsx】

即为biobert\_classifier.py用到的40个英文病例数据集（20个内科病例与20个外科病例）

The dataset of 40 English case notes (20 from Internal Medicine and 20 from Surgery) used in biobert\_classifier.py.

【biobert-local】

从hugging face上下载BioBERT为本地模型,包含：

config.json、pytorch\_model.bin、tokenizer\_config.json、vocab。而pytorch\_model.bin由于文件太大，需自行从页面<https://huggingface.co/dmis-lab/biobert-v1.1/tree/>

下载pytorch\_model.bin至本地文件夹biobert-local里，即可在本地部署BioBERT模型。

To obtain a local copy of the BioBERT model from Hugging Face, the following files should be downloaded: config.json, pytorch\_model.bin, tokenizer\_config.json, and vocab. Due to its large size, the pytorch\_model.bin file must be downloaded manually from [https://huggingface.co/dmis-lab/biobert-v1.1/tree/](https://huggingface.co/dmis-lab/biobert-v1.1/tree/" \t "_new) and placed in the local directory biobert-local to enable local deployment of the BioBERT model.