Etape 1 :

Collecte de donnes : chinois + base en anglais qu’on fera une traduction a l’aide d’un transformer MarianMT (MarianMT est une famille de modèles de traduction neuronale basés sur l'architecture Transformer, développés par le groupe Helsinki-NLP. Ils sont utilisés pour la traduction automatique entre différentes langues et sont disponibles dans la bibliothèque Hugging Face Transformers.)

Etape de classification des intentions

hfl/chinese-bert-wwm" **pour la classification d'intentions sur votre jeu de données**

"hfl/chinese-bert-wwm" est un modèle de classification. Plus précisément, c'est un modèle de classification de séquence.

Voici pourquoi :

* **Classification:** Le modèle est conçu pour attribuer une catégorie (label) à une entrée (une phrase ou un texte court dans ce cas). Dans votre exemple, les catégories sont "information", "symptômes", et "traitement".
* **Séquence:** Le modèle traite l'entrée comme une séquence de mots (une phrase en chinois et en anglais dans votre cas). Il prend en compte l'ordre et la relation entre les mots pour faire sa prédiction.
* "hfl/chinese-bert-wwm" a la capacité de prendre en compte les synonymes et la reformulation des phrases dans une certaine mesure. Voici comment :
* **1. Embeddings de mots:**
* Le modèle utilise des embeddings de mots, qui sont des vecteurs numériques représentant le sens des mots. Ces embeddings sont construits en utilisant des techniques d'apprentissage automatique qui capturent les relations sémantiques entre les mots.
* **2. Attention:**
* Le modèle utilise un mécanisme d'attention qui lui permet de se concentrer sur les parties les plus importantes de la phrase pour la classification. Cela permet au modèle de comprendre des phrases reformulées avec des mots différents mais ayant le même sens.
* **3. Limitations:**
* Cependant, il est important de noter que la capacité du modèle à comprendre les synonymes et la reformulation est limitée. Le modèle peut ne pas être en mesure de comprendre des phrases qui sont très différentes de la phrase d'apprentissage sur laquelle il a été entraîné.

**Exemple d'utilisation d'un modèle pré-entraîné pour la classification de texte**

**Introduction**

Dans cette section, nous allons explorer l'utilisation d'un modèle pré-entraîné pour la classification de texte. Nous prendrons l'exemple du modèle "hfl/chinese-bert-wwm" pour illustrer les concepts et les étapes clés.

**Modèle pré-entraîné "hfl/chinese-bert-wwm"**

Le modèle "hfl/chinese-bert-wwm" est un modèle de classification de séquence basé sur l'architecture Transformer. Il a été pré-entraîné sur un ensemble de données massif de texte chinois. Ce modèle peut être utilisé pour différentes tâches de classification de texte, comme la classification d'intention, la classification d'opinion et la classification de sujet.

**Exemple d'application**

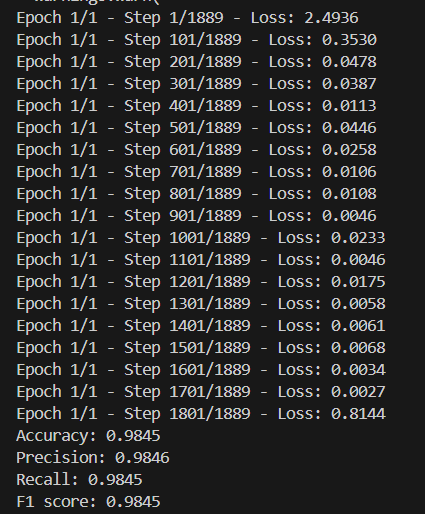
Supposons que nous voulons construire un système capable de classer les questions médicales en fonction de leur intention (information, symptômes, traitement). Nous pouvons utiliser le modèle "hfl/chinese-bert-wwm" pour cette tâche en suivant les étapes suivantes :

**1. Chargement du modèle et du tokenizer**

**2. Préparation des données**

**3. Entraînement du modèle**

**4. Évaluation du modèle**



Une image contenant texte, capture d’écran, Police, logiciel

Description générée automatiquement

Une image contenant texte, capture d’écran, Police, ligne

Description générée automatiquement

information

Do you have information about Kidney disease?

您有关于肾病的信息吗？

Greeting

Hello

您好

traduction

Can you translate this Chinese into English: What is the best way to treat diabetes?

您能将这句中文翻译成英文吗 : 治疗糖尿病的最佳方法是什么？