说 明 书 摘 要

本发明提供一种语音转录文本聚类方法、装置、电子设备和存储介质，所述方法包括：提取各语音转录文本的向量表示；将各语音转录文本的向量表示输入至文本聚类模型，得到文本聚类模型输出的各语音转录文本的聚类结果；本发明以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标训练迭代式地更新文本聚类模型，最终使得文本聚类模型能够从文本层面和类别层面对各语音转录文本进行聚类，进而准确得到聚类结果。

摘 要 附 图



权 利 要 求 书

1、一种语音转录文本聚类方法，其特征在于，包括：

提取各语音转录文本的向量表示；

将各语音转录文本的向量表示输入至文本聚类模型，得到所述文本聚类模型输出的各语音转录文本的聚类结果，并根据聚类结果更新聚类模型；

所述文本聚类模型基于多个样本语音转录文本的向量表示以及各样本语音转录文本的聚类结果训练得到，所述文本聚类模型的训练以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标。

2、根据权利要求1所述的语音转录文本聚类方法，其特征在于，所述文本聚类模型基于如下步骤训练得到：

提取各样本语音转录文本的样本向量表示，并基于各样本向量表示进行文本聚类，得到各样本语音转录文本的初始聚类结果；

基于相同类别中各样本语音转录文本的样本向量表示，确定各类别的语义向量；

基于相同样本语音转录文本的向量表示之间的距离，不同样本语音转录文本的向量表示之间的距离，样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及样本语音转录文本的向量表示与其它类别的语义向量之间的距离，确定所述文本聚类模型的损失值，并基于所述损失值对所述文本聚类模型的初始模型进行训练，更新文本聚类模型，然后重复以上步骤直至模型达到理想效果。

3、根据权利要求2所述的语音转录文本聚类方法，其特征在于，所述基于相同样本语音转录文本的向量表示之间的距离，不同样本语音转录文本的向量表示之间的距离，样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及样本语音转录文本的向量表示与其它类别的语义向量之间的距离，确定所述文本聚类模型的损失值，包括：

基于相同样本语音转录文本的向量表示之间的余弦相似度，以及不同样本语音转录文本的向量表示之间的余弦相似度，确定文本层面的对比损失值；

基于样本语音转录文本的向量表示与其所属类别的语义向量之间的余弦相似度以及样本语音转录文本的向量表示与其它类别的语义向量之间的余弦相似度，确定类别层面的对比损失值；

基于所述文本层面的对比损失值，以及所述类别层面的对比损失值，确定所述文本聚类模型的损失值。

4、根据权利要求3所述的语音转录文本聚类方法，其特征在于，所述文本层面的对比损失值基于如下公式确定：

其中，表示所述文本层面的对比损失值，表示所述相同样本语音转录文本的向量表示之间的余弦相似度，表示所述不同样本语音转录文本的向量表示之间的余弦相似度，表示余弦值的缩放程度，避免训练过程中出现梯度消失等问题，表示一个训练批次的样本数量；

所述类别层面的对比损失值基于如下公式确定：

其中，表示所述类别层面的对比损失值，表示所述样本语音转录文本的向量表示与其所属类别的语义向量之间的余弦相似度，表示所述样本语音转录文本的向量表示与其它类别的语义向量之间的余弦相似度，表示语义向量对应类别中样本语音转录文本的数量，表示语义向量对应类别中样本语音转录文本的数量，为平滑系数。

5、根据权利要求1至4任一项所述的语音转录文本聚类方法，其特征在于，所述提取各语音转录文本的向量表示，包括：

对各语音转录文本进行编码，得到各语音转录文本的字符编码向量集合；

对所述字符编码向量集合中的各向量进行均值处理，得到各语音转录文本的向量表示。

6、根据权利要求5所述的语音转录文本聚类方法，其特征在于，各语音转录文本的向量表示基于如下公式确定：

其中，表示各语音转录文本的向量表示，表示各语音转录文本的字符数，表示各语音转录文本起始字符的编码向量，表示各语音转录文本结束字符的编码向量，表示各语音转录文本的字符编码向量。

7、根据权利要求1至4任一项所述的语音转录文本聚类方法，其特征在于，所述相同样本语音转录文本的向量表示之间的距离基于如下步骤确定：

对各样本语音转录文本进行数据增强，得到各样本语音转录文本的增强文本，并提取各增强文本的向量表示；

基于各样本语音转录文本的向量表示及其对应的各增强文本的向量表示，确定所述相同样本语音转录文本的向量表示之间的距离。

8、一种语音转录文本聚类装置，其特征在于，包括：

提取单元，用于提取各语音转录文本的向量表示；

聚类单元，用于将各语音转录文本的向量表示输入至文本聚类模型，得到所述文本聚类模型输出的各语音转录文本的聚类结果；

所述文本聚类模型基于多个样本语音转录文本的向量表示以及各样本语音转录文本的聚类结果训练得到，所述文本聚类模型的训练以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标。

9、一种电子设备，包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述程序时实现如权利要求1至7任一项所述语音转录文本聚类方法。

10、一种非暂态计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求1至7任一项所述语音转录文本聚类方法。

说 明 书

# 语音转录文本聚类方法、装置、电子设备和存储介质

技术领域

本发明涉及语音转录技术领域，尤其涉及一种语音转录文本聚类方法、装置、电子设备和存储介质。

背景技术

随着自动语音识别技术（Automatic Speech Recognition，ASR）的迅速发展，出现了大量的中文ASR语音转录文本。

由于录音背景有噪声、识别技术精度有限等原因，这些语音转录文本普遍具有字错误率高、语句不流畅等特点，导致现有的文本聚类方法，如k-means等，直接应用于ASR语音转录文本时效果较差。

发明内容

本发明提供一种语音转录文本聚类方法、装置、电子设备和存储介质，用以解决现有技术中文本聚类精度较低的缺陷。

本发明提供一种语音转录文本聚类方法，包括：

提取各语音转录文本的向量表示；

将各语音转录文本的向量表示输入至文本聚类模型，得到所述文本聚类模型输出的各语音转录文本的聚类结果；

所述文本聚类模型基于多个样本语音转录文本的向量表示以及各样本语音转录文本的聚类结果训练得到，所述文本聚类模型的训练以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标。

根据本发明提供的一种语音转录文本聚类方法，所述文本聚类模型基于如下步骤训练得到：

提取各样本语音转录文本的样本向量表示，并基于各样本向量表示进行文本聚类，得到各样本语音转录文本的初始聚类结果；

基于相同类别中各样本语音转录文本的样本向量表示，确定各类别的语义向量；

基于相同样本语音转录文本的向量表示之间的距离，不同样本语音转录文本的向量表示之间的距离，样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及样本语音转录文本的向量表示与其它类别的语义向量之间的距离，确定所述文本聚类模型的损失值，并基于所述损失值对所述文本聚类模型的初始模型进行训练，更新文本聚类模型，然后重复以上步骤直至模型达到理想效果。

根据本发明提供的一种语音转录文本聚类方法，所述基于相同样本语音转录文本的向量表示之间的距离，不同样本语音转录文本的向量表示之间的距离，样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及样本语音转录文本的向量表示与其它类别的语义向量之间的距离，确定所述文本聚类模型的损失值，包括：

基于相同样本语音转录文本的向量表示之间的余弦相似度，以及不同样本语音转录文本的向量表示之间的余弦相似度，确定文本层面的对比损失值；

基于样本语音转录文本的向量表示与其所属类别的语义向量之间的余弦相似度以及样本语音转录文本的向量表示与其它类别的语义向量之间的余弦相似度，确定类别层面的对比损失值；

基于所述文本层面的对比损失值，以及所述类别层面的对比损失值，确定所述文本聚类模型的损失值。

根据本发明提供的一种语音转录文本聚类方法，所述文本层面的对比损失值基于如下公式确定：

其中，表示所述文本层面的对比损失值，表示所述相同样本语音转录文本的向量表示之间的余弦相似度，表示所述不同样本语音转录文本的向量表示之间的余弦相似度，表示余弦值的缩放程度，避免训练过程中出现梯度消失等问题，表示一个训练批次的样本数量；

所述类别层面的对比损失值基于如下公式确定：

其中，表示所述类别层面的对比损失值，表示所述样本语音转录文本的向量表示与其所属类别的语义向量之间的余弦相似度，表示所述样本语音转录文本的向量表示与其它类别的语义向量之间的余弦相似度，表示语义向量对应类别中样本语音转录文本的数量，表示语义向量对应类别中样本语音转录文本的数量，为平滑系数。

根据本发明提供的一种语音转录文本聚类方法，所述提取各语音转录文本的向量表示，包括：

对各语音转录文本进行编码，得到各语音转录文本的字符编码向量集合；

对所述字符编码向量集合中的各向量进行均值处理，得到各语音转录文本的向量表示。

根据本发明提供的一种语音转录文本聚类方法，各语音转录文本的向量表示基于如下公式确定：

其中，表示各语音转录文本的向量表示，表示各语音转录文本的字符数，表示各语音转录文本起始字符的编码向量，表示各语音转录文本结束字符的编码向量，表示各语音转录文本的字符编码向量。

根据本发明提供的一种语音转录文本聚类方法，所述相同样本语音转录文本的向量表示之间的距离基于如下步骤确定：

对各样本语音转录文本进行数据增强，得到各样本语音转录文本的增强文本，并提取各增强文本的向量表示；

基于各样本语音转录文本的向量表示及其对应的各增强文本的向量表示，确定所述相同样本语音转录文本的向量表示之间的距离。

本发明还提供一种语音转录文本聚类装置，包括：

提取单元，用于提取各语音转录文本的向量表示；

聚类单元，用于将各语音转录文本的向量表示输入至文本聚类模型，得到所述文本聚类模型输出的各语音转录文本的聚类结果；

所述文本聚类模型基于多个样本语音转录文本的向量表示以及各样本语音转录文本的聚类结果训练得到，所述文本聚类模型的训练以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标。

本发明还提供一种电子设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时实现如上述任一种所述语音转录文本聚类方法。

本发明还提供一种非暂态计算机可读存储介质，其上存储有计算机程序，该计算机程序被处理器执行时实现如上述任一种所述语音转录文本聚类方法。

本发明还提供一种计算机程序产品，包括计算机程序，所述计算机程序被处理器执行时实现如上述任一种所述语音转录文本聚类方法。

本发明提供的语音转录文本聚类方法、装置、电子设备和存储介质，以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标训练得到文本聚类模型，从而使得文本聚类模型能够从文本层面和类别层面对各语音转录文本进行聚类，进而准确得到聚类结果。

附图说明

为了更清楚地说明本发明或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1是本发明提供的语音转录文本聚类方法的流程示意图；

图2是本发明提供的文本聚类模型训练方法的流程示意图；

图3是本发明提供的语音转录文本聚类装置的结构示意图；

图4是本发明提供的电子设备的结构示意图。

具体实施方式

为使本发明的目的、技术方案和优点更加清楚，下面将结合本发明中的附图，对本发明中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

由于录音背景有噪声、识别技术精度有限等原因，这些语音转录文本普遍具有字错误率高、语句不流畅等特点，导致现有的文本聚类方法，如k-means等，直接应用于ASR语音转录文本时效果较差。

对此，本发明提供一种语音转录文本聚类方法。图1是本发明提供的语音转录文本聚类方法的流程示意图，如图1所示，该方法包括如下步骤：

步骤110、提取各语音转录文本的向量表示。

此处，语音转录文本即需要进行文本聚类的文本，该语音转录文本可以是对录音音频进行语音识别得到的文本。各语音转录文本的向量表示用于表征各语音转录文本的语义信息，其可以是对各语音转录文本进行编码获取。

步骤120将各语音转录文本的向量表示输入至文本聚类模型，得到文本聚类模型输出的各语音转录文本的聚类结果；

文本聚类模型基于多个样本语音转录文本的向量表示以及各样本语音转录文本的聚类结果训练得到，文本聚类模型的训练以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标。

具体地，相同样本语音转录文本的向量表示之间的距离用于表征两个相同样本语音转录文本之间的差异，如对任一样本语音转录文本进行数据增强后，得到增强文本，以对应样本语音转录文本和增强文本的向量表示之间的距离作为相同样本语音转录文本的向量表示之间的距离。不同样本语音转录文本的向量表示之间的距离用于表征两个不同样本语音转录文本之间的差异，如将多个样本语音转录文本中任意两个不同的样本语音转录文本的向量表示之间的距离作为不同样本语音转录文本的向量表示之间的距离。本发明实施例以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离为目标训练文本聚类模型，能够从各样本语音转录文本本身层面进行对比学习，即从文本层面学习相同样本语音转录文本之间的信息以及不同样本语音转录文本之间的信息。

此外，不同样本语音转录文本的语义信息不同，若语义信息相似度较高，则可以将对应的样本语音转录文本聚为一类，若语义相似度较低，则可以将对应的样本语音转录文本划分至两个不同的类别。对于属于同一类别的样本语音转录文本，可以基于对应样本语音转录文本的向量表示确定聚类中心，并将聚类中心作为对应类别的语义向量。样本语音转录文本的向量表示与其所属类别的语义向量之间的距离用于表征样本语音转录文本的向量表示与同类别的语义向量之间的差异，样本语音转录文本的向量表示与其它类别的语义向量之间的距离用于表示样本语音转录文本的向量表示与不同类别的语义向量之间的差异。本发明实施例以最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标训练文本聚类模型，能够从不同类别样本语音转录文本层面进行对比学习，即从类别层面学习同类样本语音转录文本之间的信息以及不同类样本语音转录文本之间的信息。

由此可见，本发明实施例提供的语音转录文本聚类方法，以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标训练得到文本聚类模型，从而使得文本聚类模型能够从文本层面和类别层面对各语音转录文本进行聚类，进而准确得到聚类结果。

基于上述实施例，文本聚类模型基于如下步骤训练得到：

提取各样本语音转录文本的样本向量表示，并基于各样本向量表示进行文本聚类，得到各样本语音转录文本的初始聚类结果；

基于相同类别中各样本语音转录文本的样本向量表示，确定各类别的语义向量；

基于相同样本语音转录文本的向量表示之间的距离，不同样本语音转录文本的向量表示之间的距离，样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及样本语音转录文本的向量表示与其它类别的语义向量之间的距离，确定文本聚类模型的损失值，并基于损失值对文本聚类模型的初始模型进行训练，更新文本聚类模型，然后重复聚类、训练的过程直至模型效果达标。

具体地，各样本语音转录文本的向量表示用于表征各样本语音转录文本的语义信息，其可以是对各样本语音转录文本进行编码获取。基于各样本向量表示，可以确定各样本向量表示之间的距离，距离越大，表明对应样本语音转录文本之间的差异越大，也就是对应样本语音转录文本属于不同类别的概率越大。反之，距离越小，表明对应样本语音转录文本之间的差异越小，也就是对应样本语音转录文本属于相同类别的概率越大。基于此，基于各样本向量表示进行文本聚类，可以得到各样本语音转录文本的初始聚类结果，可选地，可以采用K均值聚类算法对各样本向量表示进行聚类，得到初始聚类结果。

在得到初始聚类结果后，即可得到各类别中包含的样本语音转录文本，进而基于各类别中样本语音转录文本的样本向量表示，确定各类别的中心，并以各类别的中心作为各类别的语义向量。

在得到各类别的语义向量后，基于相同样本语音转录文本的向量表示之间的距离，不同样本语音转录文本的向量表示之间的距离，样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及样本语音转录文本的向量表示与其它类别的语义向量之间的距离，确定文本聚类模型的损失值，进而基于损失值对文本聚类模型的初始模型进行训练，从而可以使得得到的文本聚类模型能够从文本层面和类别层面对各语音转录文本进行聚类，进而准确得到聚类结果。

需要说明的是，在每次基于损失值对文本聚类模型的初始模型进行训练后\*\*\*

基于上述任一实施例，基于相同样本语音转录文本的向量表示之间的距离，不同样本语音转录文本的向量表示之间的距离，样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及样本语音转录文本的向量表示与其它类别的语义向量之间的距离，确定文本聚类模型的损失值，包括：

基于相同样本语音转录文本的向量表示之间的余弦相似度，以及不同样本语音转录文本的向量表示之间的余弦相似度，确定文本层面的对比损失值；

基于样本语音转录文本的向量表示与其所属类别的语义向量之间的余弦相似度以及样本语音转录文本的向量表示与其它类别的语义向量之间的余弦相似度，确定类别层面的对比损失值；

基于文本层面的对比损失值，以及类别层面的对比损失值，确定文本聚类模型的损失值。

具体地，相同样本语音转录文本的向量表示之间的余弦相似度用于表征相同样本语音转录文本的向量表示之间的距离，不同样本语音转录文本的向量表示之间的余弦相似度用于表征不同样本语音转录文本的向量表示之间的距离，基于两者可以得到文本层面的对比损失值。

此外，样本语音转录文本的向量表示与其所属类别的语义向量之间的余弦相似度用于表征样本语音转录文本的向量表示与其所属类别的语义向量之间的距离，样本语音转录文本的向量表示与其它类别的语义向量之间的余弦相似度用于表征样本语音转录文本的向量表示与其它类别的语义向量之间的距离，基于两者可以得到类别层面的对比损失值。

最后，基于文本层面的对比损失值，以及类别层面的对比损失值，确定文本聚类模型的损失值，如对文本层面的对比损失值和类别层面的对比损失值进行权重相加，得到文本聚类模型的损失值。

基于上述任一实施例，文本层面的对比损失值基于如下公式确定：

其中，表示文本层面的对比损失值，表示相同样本语音转录文本的向量表示之间的余弦相似度，表示不同样本语音转录文本的向量表示之间的余弦相似度，表示余弦值的缩放程度，避免训练过程中出现梯度消失等问题，表示一个训练批次的样本数量。其中，文本层面的对比损失值的目的是为了使得各样本语音转录文本与基于其自身生成的正样本在特征空间中更近，与其他不同样本语音转录文本在特征空间中更远。

类别层面的对比损失值基于如下公式确定：

其中，表示类别层面的对比损失值，表示样本语音转录文本的向量表示与其所属类别的语义向量之间的余弦相似度，表示样本语音转录文本的向量表示与其它类别的语义向量之间的余弦相似度，表示语义向量对应类别中样本语音转录文本的数量，表示语义向量对应类别中样本语音转录文本的数量，为平滑系数，防止聚类松散程度趋于正无穷。和表示聚类松散程度，用于衡量对应类别的语义向量的真实性，具体来说，聚类松散程度越小，则说明该类别的文本在空间中分布越集中，越认为对应类别的语义向量可以很好地代表该类别的语义。

其中，聚类松散程度越小，对应类别的语义向量提供的损失也就越多，类别层面的对比损失值的目标是使得样本语音转录文本与其聚类所属类别的语义向量在特征空间中更近，与其他类别的语义向量在特征空间中更远，从而使算法更完美地完成按语义对文本进行聚类的目标。

可选地，在得到文本层面的对比损失值，以及类别层面的对比损失值后，可以基于如下公式确定文本聚类模型的损失值：

其中，表示文本聚类模型的损失值，是平衡文本层面的对比损失值和类别层面的对比损失值的参数，最后可以使用Adam（一种反向传播算法的变体）来优化损失函数，训练用于获取文本表示的编码器，以获得可训练参数的值。

基于上述任一实施例，提取各语音转录文本的向量表示，包括：

对各语音转录文本进行编码，得到各语音转录文本的字符编码向量集合；

对字符编码向量集合中的各向量进行均值处理，得到各语音转录文本的向量表示。

具体地，在对各语音转录文本进行编码时，可以采用Bert/Roberta等预训练语言模型获取各语音转录文本的字符编码向量集合，集合中的每个向量表示语音转录文本中的字符经过编码之后的字符向量。通常情况下，使用作为语音转录文本的向量表示，但是在聚类任务中，因为下游不会连接任何分类层，聚类时无法有效捕捉语义信息，所以本发明实施例放弃使用，而是对文本编码后的字符编码向量集合中的所有向量作均值处理（即均值池化处理），提取各语音转录文本的向量表示。

需要说明的是，由于语音转录文本的语义信息，例如情感倾向，通常是由语音转录文本中大部分字符的情感倾向共同决定的，所以本发明实施例使用均值池化方法来提取语音转录文本的向量表示，使语料库中具有相似语义内容的文本可以在向量空间中有着更相似的文本表示。此外，均值池化层可以有效减轻语音转录文本中少量的错误字词对分类结果的影响，从而能够进一步提高聚类结果的精度。

基于上述任一实施例，各语音转录文本的向量表示基于如下公式确定：

其中，表示各语音转录文本的向量表示，表示各语音转录文本的字符数，表示各语音转录文本起始字符的编码向量，表示各语音转录文本结束字符的编码向量，表示各语音转录文本的字符编码向量。

基于上述任一实施例，相同样本语音转录文本的向量表示之间的距离基于如下步骤确定：

对各样本语音转录文本进行数据增强，得到各样本语音转录文本的增强文本，并提取各增强文本的向量表示；

基于各样本语音转录文本的向量表示及其对应的各增强文本的向量表示，确定相同样本语音转录文本的向量表示之间的距离。

具体地，利用深度学习模型dropout机制的随机性，对同一个样本语音转录文本再次进行编码，得到该文本的另一个向量表示，也即该增强文本的向量表示，并将其与对应样本语音转录文本的向量表示作为一个正样本对，并确定该正样本对中两个向量表示之间的距离，即为相同样本语音转录文本的向量表示之间的距离。

基于上述任一实施例，本发明还提供一种文本聚类模型的训练方法，如图2所示，该方法包括如下步骤：

首先，收集样本语音转录文本并进行数据预处理。接着，采用编码器提取样本语音转录文本的字符编码向量集合，并对字符编码向量集合中的所有向量进行均值处理，得到样本语音转录文本的向量表示。其中，编码器可以是基于预训练的Bert模型训练得到的。

随即，基于样本语音转录文本的向量表示，采用K均值聚类算法对样本语音转录文本进行聚类，得到初始聚类结果。在得到初始聚类结果后，基于相同样本语音转录文本的向量表示之间的余弦相似度，以及不同样本语音转录文本的向量表示之间的余弦相似度，确定文本层面的对比损失值，基于样本语音转录文本的向量表示与其所属类别的语义向量之间的余弦相似度以及样本语音转录文本的向量表示与其它类别的语义向量之间的余弦相似度，确定类别层面的对比损失值；然后，基于文本层面的对比损失值，以及类别层面的对比损失值，确定文本聚类模型的损失值。

接着，基于文本聚类模型的损失值进行反向传播，以对初始模型进行更新，更新完毕后重新进行聚类、模型训练、模型更新，最终得到训练完成的文本聚类模型。

下面对本发明提供的语音转录文本聚类装置进行描述，下文描述的语音转录文本聚类装置与上文描述的语音转录文本聚类方法可相互对应参照。

基于上述任一实施例，本发明还提供一种语音转录文本聚类装置，如图3所示，该装置包括：

提取单元310，用于提取各语音转录文本的向量表示；

聚类单元320，用于将各语音转录文本的向量表示输入至文本聚类模型，得到所述文本聚类模型输出的各语音转录文本的聚类结果；

所述文本聚类模型基于多个样本语音转录文本的向量表示以及各样本语音转录文本的聚类结果训练得到，所述文本聚类模型的训练以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标。

图4是本发明提供的电子设备的结构示意图，如图4所示，该电子设备可以包括：处理器(processor)410、存储器(memory)420、通信接口(Communications Interface)430和通信总线440，其中，处理器410，存储器420，通信接口430通过通信总线440完成相互间的通信。处理器410可以调用存储器420中的逻辑指令，以执行语音转录文本聚类方法，该方法包括：提取各语音转录文本的向量表示；将各语音转录文本的向量表示输入至文本聚类模型，得到所述文本聚类模型输出的各语音转录文本的聚类结果；所述文本聚类模型基于多个样本语音转录文本的向量表示以及各样本语音转录文本的聚类结果训练得到，所述文本聚类模型的训练以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标。

此外，上述的存储器420中的逻辑指令可以通过软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

另一方面，本发明还提供一种计算机程序产品，所述计算机程序产品包括存储在非暂态计算机可读存储介质上的计算机程序，所述计算机程序包括程序指令，当所述程序指令被计算机执行时，计算机能够执行上述各方法所提供的语音转录文本聚类方法，该方法包括：提取各语音转录文本的向量表示；将各语音转录文本的向量表示输入至文本聚类模型，得到所述文本聚类模型输出的各语音转录文本的聚类结果；所述文本聚类模型基于多个样本语音转录文本的向量表示以及各样本语音转录文本的聚类结果训练得到，所述文本聚类模型的训练以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标。

又一方面，本发明还提供一种非暂态计算机可读存储介质，其上存储有计算机程序，该计算机程序被处理器执行时实现以执行上述各提供的语音转录文本聚类方法，该方法包括：提取各语音转录文本的向量表示；将各语音转录文本的向量表示输入至文本聚类模型，得到所述文本聚类模型输出的各语音转录文本的聚类结果；所述文本聚类模型基于多个样本语音转录文本的向量表示以及各样本语音转录文本的聚类结果训练得到，所述文本聚类模型的训练以最小化相同样本语音转录文本的向量表示之间的距离，最大化不同样本语音转录文本的向量表示之间的距离，最小化样本语音转录文本的向量表示与其所属类别的语义向量之间的距离以及最大化样本语音转录文本的向量表示与其它类别的语义向量之间的距离为目标。

以上所描述的装置实施例仅仅是示意性的，其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下，即可以理解并实施。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件。基于这样的理解，上述技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在计算机可读存储介质中，如ROM/RAM、磁碟、光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行各个实施例或者实施例的某些部分所述的方法。

最后应说明的是：以上实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

说 明 书 附 图



图1



图2



图3



图4