摘要

人体动作的生成对机器人工程、游戏或动画等诸多应用领域具有价值。最近，为了节约费用和时间，自动动作生成方法很受欢迎，而文本可以比较好的描述人体动作，所以基于文本描述的动作生成方法对研究者来说是非常重要的研究课题。然而，同时理解文本描述和人体动作的基于文本描述的人体动作生成是一项非常艰巨的任务。近几年，虽然一些基于文本的人体动作生成方法被提出了，但是数据集和方法方面仍然存在问题和局限性。首先，现数据集方面存在数据量严重不足、动作分布不均匀等问题。然后，方法方面比较注重在数据集分布内的动作生成上，在数据集分布外的动作生成上的效果不佳，导致无法生成很多现实世界中的动作。因此，本文对数据集和方法所存在的问题进行分析与研究，以改善问题和局限性，从而生成出更多来自现实世界中的动作。本文主要内容如下：

（1）本文提出了动作与文本描述新数据集，包括大约一万六千对的丰富的动作数据和文本描述。该数据集包含了比现有数据集多 5 倍的动作数据和多 2 倍的文本描述，动作数据分布更均匀。本文使用了现有的基于文本的动作生成方法进行了实验和分析，通过实验验证了与现有数据集相比，该数据集可以更客观地评估生成结果，相比使用现有数据集，可以生成更多的来自现实世界中的动作。

（2）本文提出了一个新的框架，它利用了大型语言模型的知识和常识处理能力，旨在处理数据集分布外的复杂动作文本或多种动作组合而成的长文本。该框架由三个阶段组成，在第一个阶段，本文使用了大型语言模型将复杂的文本分解为易于理解的阶段文本，在第二阶段，本文基于分解后的文本生成动作，在第三阶段，本文将所有生成的动作合成一个完整、自然的动作。本文通过实验对比，验证了可以较好地生成数据集分布外的复杂动作和多个动作组成合而成的长序列动作。

综上所述，本文提出的一个新的数据集和一个新的框架，在一定的程度上克服了基于文本描述的人体动作生成研究面临的问题和局限性，为该研究领域提供了有效的支持。

关键词：动作生成，人体动作，深度学习

Research on Text Description-based Human Motion Generation

RO,DONGWOO (Computer Applied Technology)
Directed by: Prof. Yizhou Wang

ABSTRACT

Human motion generation has many applications such as robotics, games, and animation. Recently, in order to decrease costs and time, the automatic motion generation method was an often implemented approach, and text was able to describe human motion in detail, so the text-based motion generation method is a significant research topic for researchers. However, it is a very difficult task to understand both the textual description and human motion. In recent years, although some text-based methods for generating human motions were proposed, there are still some challenges and limitations in datasets and methods. The first challenge is that the motion data in the dataset are insufficient, and they also have uneven data distribution; a particular motion category takes up the majority of the entire dataset. Also, the methods more focus on motion generation within the dataset distribution, and it can fail to generate motion in the out-of-distribution of datasets, which results in failure to generate sufficient motions from the real world. This paper analyzed the existing challenges of datasets and methods to overcome the problems and the limitations and to generate more motions in the real world.

The main contributions are as follows:

(1) This paper presented a new dataset of motion and text descriptions, including about sixteen thousand pairs of motion data and text descriptions. The dataset contains 5 times more motion data and 2 times more text descriptions than the existing dataset, and the motion data is more evenly distributed. Experiments and analysis were carried out using the existing text-based motion generation method. Compared with the existing dataset, the dataset could evaluate the result objectively. Compared with using the existing dataset, it could generate more motions in the real world.

(2) A new framework was proposed, which utilizes the knowledge and commonsense processing capabilities of Large Language Models to deal with complex motion text or long text composed of multiple actions out-of-distribution of the dataset. In the first stage, this paper used a Large Language Model to decompose a complex text into easily understood text series. In the second stage, this paper generated motions based on the decomposed text series. In the

last stage, this paper combined motion sequences into one completed and natural motion. The experimental comparison shows that the complex motions and the long sequence motions can be generated well which is out-of-distribution of the dataset.

In conclusion, a new dataset and a new framework were proposed to overcome the problems and limitations of text-based human motion generation research and provide adequate support for the research field.

KEY WORDS: Motion Generation, Human Motion, Deep Learning