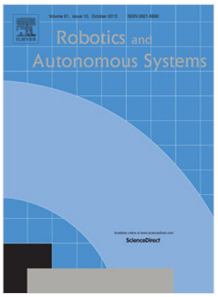
Accepted Manuscript



Learning a bidirectional mapping between human whole-body motion and natural language using deep recurrent neural networks

Matthias Plappert, Christian Mandery, Tamim Asfour

PII: S0921-8890(17)30628-0
DOI: https://doi.org/10.1016/j.robot.2018.07.006
Reference: ROBOT 3060

To appear in: *Robotics and Autonomous Systems*

Received date : 14 September 2017
Revised date : 5 June 2018
Accepted date : 24 July 2018

Please cite this article as: M. Plappert, C. Mandery, T. Asfour, Learning a bidirectional mapping between human whole-body motion and natural language using deep recurrent neural networks, *Robotics and Autonomous Systems* (2018), https://doi.org/10.1016/j.robot.2018.07.006

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Learning a bidirectional mapping between human whole-body motion and natural language using deep recurrent neural networks

Matthias Plappert*, Christian Mandery, Tamim Asfour

High Performance Humanoid Technologies (H^2T), Karlsruhe Institute of Technology (KIT), Adenauerring 2 (Building 50.20), 76131 Karlsruhe, Germany

Abstract

Linking human whole-body motion and natural language is of great interest for the generation of semantic representations of observed human behaviors as well as for the generation of robot behaviors based on natural language input. While there has been a large body of research in this area, most approaches that exist today require a symbolic representation of motions (e.g. in the form of motion primitives), which have to be defined a-priori or require complex segmentation algorithms. In contrast, recent advances in the field of neural networks and especially deep learning have demonstrated that sub-symbolic representations that can be learned end-to-end usually outperform more traditional approaches, for applications such as machine translation. In this paper we propose a generative model that learns a bidirectional mapping between human whole-body motion and natural language using *deep recurrent neural networks* (RNNs) and *sequence-to-sequence learning*. Our approach does not require any segmentation or manual feature engineering and learns a distributed representation, which is shared for all motions and descriptions. We evaluate our approach on 2 846 human whole-body motions and 6 187 natural language descriptions thereof from the *KIT Motion-Language Dataset*. Our results clearly demonstrate the effectiveness of the proposed model: We show that our model generates a wide variety of realistic motions only from descriptions thereof in form of a single sentence. Conversely, our model is also capable of generating correct and detailed natural language descriptions from human motions.

Keywords: human whole-body motion; natural language; sequence-to-sequence learning; recurrent neural network

Download English Version:

https://daneshyari.com/en/article/8960159

Download Persian Version:

https://daneshyari.com/article/8960159

Daneshyari.com