Accepted Manuscript

Character-Level Neural Network for Biomedical Named Entity Recognition

Mourad Gridach

Please cite this article as: Gridach, M., Character-Level Neural Network for Biomedical Named Entity Recognition, *Journal of Biomedical Informatics* (2017), doi: http://dx.doi.org/10.1016/j.jbi.2017.05.002

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.

Character-Level Neural Network for Biomedical Named Entity Recognition

Mourad Gridach

High Institute of Technology, Ibn Zohr University, Agadir

Abstract

Biomedical named entity recognition (BNER), which extracts important named entities such as genes and proteins, is a challenging task in automated systems that mine knowledge in biomedical texts. The previous state-of-the-art systems required large amounts of task-specific knowledge in the form of feature engineering, lexicons and data pre-processing to achieve high performance. In this paper, we introduce a novel neural network architecture that benefits from both word- and character-level representations automatically, by using a combination of bidirectional long short-term memory (LSTM) and conditional random field (CRF) eliminating the need for most feature engineering tasks. We evaluate our system on two datasets: JNLPBA corpus and the BioCreAtIvE II Gene Mention (GM) corpus. We obtained state-of-the-art performance by outperforming the previous systems. To the best of our knowledge, we are the first to investigate the combination of deep neural networks, CRF, word embeddings and character-level representation in recognising biomedical named entities.

Keywords: Deep neural networks, natural language processing, biomedical named entity recognition

1. Introduction

Biomedical Named Entity Recognition (BNER) is a critical task for extracting patient information from biomedical texts to support biomedical and trans-

Email address: m.gridach@uiz.ac.ma (Mourad Gridach)