First Workshop on Subword and Character Level Models in NLP 2017

Copenhagen, Denmark
7 September 2017

ISBN: 978-1-5108-4748-4

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2017) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2017)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Character and Subword-Based Word Representation for Neural Language Modeling Prediction
Matthieu Labeau and Alexandre Allauzen 1

Learning variable length units for SMT between related languages via Byte Pair Encoding
Anoop Kunchukuttan and Pushpak Bhattacharyya 14

Character Based Pattern Mining for Neology Detection
Gaël Lejeune and Emmanuel Cartier 25

Automated Word Stress Detection in Russian
Maria Ponomareva, Kirill Milintsevich, Ekaterina Chernyak and Anatoly Starostin 31

A Syllable-based Technique for Word Embeddings of Korean Words
Sanghyuk Choi, Taeuk Kim, Jinseok Seol and Sang-goo Lee 36

Supersense Tagging with a Combination of Character, Subword, and Word-level Representations
Youhyun Shin and Sang-goo Lee 41

Weakly supervised learning of allomorphy
Miikka Silfverberg and Mans Hulden 46

Character-based recurrent neural networks for morphological relational reasoning
Olof Mogren and Richard Johansson 57

Glyph-aware Embedding of Chinese Characters
Falcon Dai and Zheng Cai 64

Exploring Cross-Lingual Transfer of Morphological Knowledge In Sequence-to-Sequence Models
Huiming Jin and Katharina Kann 70

Unlabeled Data for Morphological Generation With Character-Based Sequence-to-Sequence Models
Katharina Kann and Hinrich Schütze 76

Vowel and Consonant Classification through Spectral Decomposition
Patricia Thaine and Gerald Penn 82

Syllable-level Neural Language Model for Agglutinative Language
Seunghak Yu, Nilesh Kulkarni, Haejun Lee and Jihie Kim 92

Character-based Bidirectional LSTM-CRF with words and characters for Japanese Named Entity Recognition
Shotaro Misawa, Motoki Taniguchi, Yasuhide Miura and Tomoko Ohkuma 97

Word Representation Models for Morphologically Rich Languages in Neural Machine Translation
Ekaterina Vylomova, Trevor Cohn, Xuanli He and Gholamreza Haffari 103

Spell-Checking based on Syllabification and Character-level Graphs for a Peruvian Agglutinative Language
Carlo Alva and Arturo Oncevay 109

What do we need to know about an unknown word when parsing German
Bich-Ngoc Do, Ines Rehbein and Anette Frank 117

A General-Purpose Tagger with Convolutional Neural Networks
Xiang Yu, Agnieszka Falenska and Ngoc Thang Vu 124

Reconstruction of Word Embeddings from Sub-Word Parameters
Karl Stratos 130

Inflection Generation for Spanish Verbs using Supervised Learning
Cristina Barros, Dimitra Gkatzia and Elena Lloret 136

Neural Paraphrase Identification of Questions with Noisy Pretraining
Gaurav Singh Tomar, Thyago Duque, Oscar Täckström, Jakob Uszkoreit and Dipanjan Das . . . 142

Sub-character Neural Language Modelling in Japanese
Viet Nguyen, Julian Brooke and Timothy Baldwin 148

Byte-based Neural Machine Translation
Marta R. Costa-jussà, Carlos Escolano and José A. R. Fonollosa 154

Improving Opinion-Target Extraction with Character-Level Word Embeddings
Soufian Jebbara and Philipp Cimiano 159