

Communications in Computer and Information Science

1215

Commenced Publication in 2007

Founding and Former Series Editors:

Simone Diniz Junqueira Barbosa, Phoebe Chen, Alfredo Cuzzocrea,
Xiaoyong Du, Orhun Kara, Ting Liu, Krishna M. Sivalingam,
Dominik Ślęzak, Takashi Washio, Xiaokang Yang, and Junsong Yuan

Editorial Board Members

Joaquim Filipe 


Polytechnic Institute of Setúbal, Setúbal, Portugal

Ashish Ghosh

Indian Statistical Institute, Kolkata, India

Igor Kotenko 

*St. Petersburg Institute for Informatics and Automation of the Russian
Academy of Sciences, St. Petersburg, Russia*

Raquel Oliveira Prates 

Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

Lizhu Zhou

Tsinghua University, Beijing, China

More information about this series at <http://www.springer.com/series/7899>

Le-Minh Nguyen · Xuan-Hieu Phan ·
Kôiti Hasida · Satoshi Tojo (Eds.)

Computational Linguistics

16th International Conference of the Pacific Association
for Computational Linguistics, PACLING 2019
Hanoi, Vietnam, October 11–13, 2019
Revised Selected Papers

Editors

Le-Minh Nguyen
Japan Advanced Institute of Science
and Technology
Ishikawa, Japan

Kôiti Hasida
Graduate School of Information Science
and Technology
The University of Tokyo
Tokyo, Japan

Xuan-Hieu Phan
University of Engineering and Technology
Hanoi, Vietnam

Satoshi Tojo
Japan Advanced Institute of Science
and Technology
Ishikawa, Japan

ISSN 1865-0929

ISSN 1865-0937 (electronic)

Communications in Computer and Information Science

ISBN 978-981-15-6167-2

ISBN 978-981-15-6168-9 (eBook)

<https://doi.org/10.1007/978-981-15-6168-9>

© Springer Nature Singapore Pte Ltd. 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This book constitutes the refereed proceedings of the 16th International Conference of the Pacific Association for Computational Linguistics (PACLING 2019), held in Hanoi, Vietnam, in October 2019. The 28 revised full papers and 14 short papers presented were carefully reviewed and selected from 70 submissions. The papers are organized into topical sections, including: Text Summarization; Relation and Word Embedding; Statistical Machine Translation; Text Classification; Web Analyzing; Question and Answering; Dialog Analyzing; Speech and Emotion Analyzing; Parsing and Segmentation; Information Extraction; Grammar Error; and Plagiarism Detection.

PACLING 2019 enjoyed many international participants, representing Vietnam, Japan, India, Canada, Taiwan, Thailand, Tunisia, and the UK. The conference is notable in that it included a session for discussing future directions and issues of computational linguistics in Vietnam. We would like to thank all the participants and the committee members for their contributions to PACLING 2019. Our gratitude also goes to the static members of FPT University for the local arrangements and hospitality. We would like to thank keynote speakers (Professor Kentaro Inui, Professor Danushka Bollegala, Dr. Kentaro Torisawa, and Professor Tomoko Matsui) for their interesting and great talks in PACLING 2019. PACLING 2019 was the 16th in the series of biannual meetings that started in 1989. The first two of these events were the Japan-Australia Joint Symposium on NLP held in Australia and then in Japan, followed by 13 PACLING conferences held not only in Australia and Japan, but also in Canada, Malaysia, Indonesia, Myanmar, and Vietnam.

February 2020

Le-Minh Nguyen
Xuan-Hieu Phan
Kôiti Hasida
Satoshi Tojo

Organization

General Chairs

Koiti Hasida
Satoshi Tojo
Nguyen Khac Thanh

The University of Tokyo, Japan
JAIST, Japan
FPT University, Vietnam

Program Chairs

Le Minh Nguyen
Xuan Hieu Phan

JAIST, Japan
VNU UET, Vietnam

Organizing Committee

Kim Anh Nguyen
(Co-chair)
Thi Minh Huyen Nguyen
(Co-chair)
Hong Viet Le
Hung Quy Pham
The Trung Tran

FPT University, Vietnam
VNU HUS, Vietnam
FPT Corporation, Vietnam
FPT University, Vietnam
FPT University, Vietnam

Program Committee

Thomas Ahmad
Kenji Araki
Vataya Chunwijitra
Kohji Dohsaka
Alexander Gelbukh
Choochart Haruechaiyasak
Koiti Hasida
Yoshihiko Hayashi
Kai Ishikawa

University of Minho, Portugal
Hokkaido University, Japan
NECTEC, Thailand
Akita Prefectural University, Japan
Instituto Politécnico Nacional, Mexico
NECTEC, Thailand
AIST, Japan
Waseda University, Japan
Data Science Research Laboratories - NEC
Corporation, Japan
Tokyo University of Technology, Japan
Shizuoka University, Japan
Dalhousie University, Canada
Kanazawa Institute of Technology, Japan
Ton Duc Thang University, Vietnam
Hanoi University of Science and Technology, Vietnam
Hanoi University of Science and Technology, Vietnam
FPT University, Vietnam

Diego Molla	Macquarie University, Australia
Huy-Tien Nguyen	JAIST, Japan
Kiem-Hieu Nguyen	Hanoi University of Science and Technology, Vietnam
Kim Anh Nguyen	Institute of Big Data - Vingroup, Vietnam
Le-Minh Nguyen	JAIST, Japan
Ngan Nguyen	Vietnam National University, Vietnam
Thai Phuong Nguyen	Vietnam National University, Vietnam
Thi Minh Huyen Nguyen	Vietnam National University, Vietnam
Tri Thanh Nguyen	Vietnam National University, Vietnam
Truong-Son Nguyen	Vietnam National University, Vietnam
Tien Nguyen-Minh	Hung Yen Technical University, Vietnam
Minh Quang Nhat Pham	Alt Vietnam Co., Ltd.
Anh Phan	Le Quy Don Technical University, Vietnam
Xuan-Hieu Phan	Vietnam National University, Vietnam
Hiroaki Saito	Keio University, Japan
Kazutaka Shimada	Kyushu Institute of Technology, Japan
Akira Shimazu	JAIST, Japan
Kiyoaki Shirai	JAIST, Japan
Thepchai Supnithi	NECTEC, Thailand
Masami Suzuki	KDDI Research, Inc., Japan
Kumiko Tanaka-Ishii	The University of Tokyo, Japan
Satoshi Tojo	JAIST, Japan
Takenobu Tokunaga	TITECH, Japan
Oanh Tran	Vietnam National University, Vietnam
Vu Tran	JAIST, Japan
Hai-Long Trieu	AIST, Japan
Tran Van Khanh	JAIST, Japan
Vinh Van Nguyen	Vietnam National University, Vietnam
Chai Wutiwiwatchai	NECTEC, Thailand
Yang Xiang	University of Guelph, Canada
Ngo Xuan Bach	Posts and Telecommunications Institute of Technology, Vietnam



Contents

Text Summarization

A Submodular Approach for Reference Recommendation.	3
<i>Thanh-Binh Kieu, Son Bao Pham, Xuan-Hieu Phan, and Massimo Piccardi</i>	
Split First and Then Rephrase: Hierarchical Generation for Sentence Simplification.	15
<i>Mengru Wang, Hiroaki Ozaki, Yuta Koreeda, and Kohsuke Yanai</i>	
Abstractive Text Summarization Using LSTMs with Rich Features	28
<i>Viet Nguyen Quoc, Huong Le Thanh, and Tuan Luu Minh</i>	

Relation and Word Embedding

SemSeq: A Regime for Training Widely-Applicable Word-Sequence Encoders.	43
<i>Hiroaki Tsuyuki, Tetsuji Ogawa, Tetsunori Kobayashi, and Yoshihiko Hayashi</i>	
Learning to Compose Relational Embeddings in Knowledge Graphs	56
<i>Wenye Chen, Huda Hakami, and Danushka Bollegala</i>	
Context-Guided Self-supervised Relation Embeddings	67
<i>Huda Hakami and Danushka Bollegala</i>	
Evaluation of Embedded Vectors for Lexemes and Synsets Toward Expansion of Japanese WordNet.	79
<i>Daiki Ko and Koichi Takeuchi</i>	
Neural Rasch Model: How Do Word Embeddings Adjust Word Difficulty?.	88
<i>Yo Ehara</i>	

Machine Translation

Dynamic Fusion: Attentional Language Model for Neural Machine Translation	99
<i>Michiki Kurosawa and Mamoru Komachi</i>	

Improving Context-Aware Neural Machine Translation with Target-Side Context	112
<i>Hayashide Yamagishi and Mamoru Komachi</i>	
Learning to Evaluate Neural Language Models	123
<i>James O'Neill and Danushka Bollegala</i>	
Recommending the Workflow of Vietnamese Sign Language Translation via a Comparison of Several Classification Algorithms	134
<i>Luyl-Da Quach, Nghia Duong-Trung, Anh-Van Vu, and Chi-Ngon Nguyen</i>	
Text Classification	
Document Classification by Word Embeddings of BERT	145
<i>Hirotaka Tanaka, Hiroyuki Shinnou, Rui Cao, Jing Bai, and Wen Ma</i>	
Deep Domain Adaptation for Low-Resource Cross-Lingual Text Classification Tasks	155
<i>Guan-Yuan Chen and Von-Wun Soo</i>	
Multi-task Learning for Aspect and Polarity Recognition on Vietnamese Datasets	169
<i>Dang Van Thin, Duc-Vu Nguyen, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, and Anh Hoang-Tu Nguyen</i>	
Evaluating Classification Algorithms for Recognizing Figurative Expressions in Japanese Literary Texts	181
<i>Mateusz Babieno, Rafal Rzepka, and Kenji Araki</i>	
Web Analysing	
Model-Driven Web Page Segmentation for Non Visual Access	191
<i>Judith Jeyafreeda Andrew, Stéphane Ferrari, Fabrice Maurel, Gaël Dias, and Emmanuel Giguet</i>	
Update Frequency and Background Corpus Selection in Dynamic TF-IDF Models for First Story Detection	206
<i>Fei Wang, Robert J. Ross, and John D. Kelleher</i>	
A Pilot Study on Argument Simplification in Stance-Based Opinions	218
<i>Pavithra Rajendran, Danushka Bollegala, and Simon Parsons</i>	
Automatic Approval of Online Comments with Multiple-Encoder Networks	231
<i>Vu Dang</i>	

Question and Answering, Dialog Analyzing

Is the Simplest Chatbot Effective in English Writing Learning Assistance? . . .	245
<i>Ryo Nagata, Tomoya Hashiguchi, and Driss Sadoun</i>	
Towards Task-Oriented Dialogue in Mixed Domains	257
<i>Tho Chi Luong and Phuong Le-Hong</i>	
Timing Prediction of Facilitating Utterance in Multi-party Conversation. . . .	267
<i>Tomonobu Sembokuya and Kazutaka Shimada</i>	
Evaluating Co-reference Chains Based Conversation History in Conversational Question Answering.	280
<i>Angrosh Mandya, Danushka Bollegala, and Frans Coenen</i>	

Speech and Emotion Analyzing

Multiple Linear Regression of Combined Pronunciation Ease and Accuracy Index.	295
<i>Katsunori Kotani and Takehiko Yoshimi</i>	
Rap Lyrics Generation Using Vowel GAN.	307
<i>Tomoya Miyano and Hiroaki Saito</i>	
Emotion Recognition for Vietnamese Social Media Text	319
<i>Vong Anh Ho, Duong Huynh-Cong Nguyen, Danh Hoang Nguyen, Linh Thi-Van Pham, Duc-Vu Nguyen, Kiet Van Nguyen, and Ngan Luu-Thuy Nguyen</i>	
Effects of Soft-Masking Function on Spectrogram-Based Instrument - Vocal Separation.	334
<i>Duc Chung Tran and M. K. A. Ahamed Khan</i>	

Parsing and Segmentation

Japanese Predicate Argument Structure Analysis with Pointer Networks. . . .	347
<i>Keigo Takahashi, Hikaru Omori, and Mamoru Komachi</i>	
An Experimental Study on Constituency Parsing for Vietnamese.	360
<i>Luong Nguyen-Thi and Phuong Le-Hong</i>	
Antonyms-Synonyms Discrimination Based on Exploiting Rich Vietnamese Features	374
<i>Bui Van Tan, Nguyen Phuong Thai, Pham Van Lam, and Dinh Khac Quy</i>	
Towards a UMLS-Integratable Vietnamese Medical Terminology	388
<i>The Quyen Ngo, My Linh Ha, Thi Minh Huyen Nguyen, Thi Mai Huong Hoang, and Viet Hung Nguyen</i>	

Vietnamese Word Segmentation with SVM: Ambiguity Reduction and Suffix Capture	400
<i>Duc-Vu Nguyen, Dang Van Thin, Kiet Van Nguyen, and Ngan Luu-Thuy Nguyen</i>	
An Assessment of Substitute Words in the Context of Academic Writing Proposed by Pre-trained and Specific Word Embedding Models	414
<i>Chooi Ling Goh and Yves Lepage</i>	
Effective Approach to Joint Training of POS Tagging and Dependency Parsing Models	428
<i>Xuan-Dung Doan, Tu-Anh Tran, and Le-Minh Nguyen</i>	
Information Extraction	
Towards Computing Inferences from English News Headlines	439
<i>Elizabeth Jasmi George and Radhika Mamidi</i>	
Extraction of Food Product and Shop Names from Blog Articles Using Named Entity Recognition	454
<i>Ryuya Ikeda and Kazuaki Ando</i>	
Transfer Learning for Information Extraction with Limited Data	469
<i>Minh-Tien Nguyen, Viet-Anh Phan, Le Thai Linh, Nguyen Hong Son, Le Tien Dung, Miku Hirano, and Hajime Hotta</i>	
Self-deprecating Humor Detection: A Machine Learning Approach	483
<i>Ashraf Kamal and Muhammad Abulaish</i>	
Grammar Error and Plagiarism Detection	
Deep Learning Approach for Vietnamese Consonant Misspell Correction . . .	497
<i>Ha Thanh Nguyen, Tran Binh Dang, and Le Minh Nguyen</i>	
Grammatical Error Correction for Vietnamese Using Machine Translation . . .	505
<i>Nghia Luan Pham, Tien Ha Nguyen, and Van Vinh Nguyen</i>	
Developing a Framework for a Thai Plagiarism Corpus	513
<i>Santipong Thaiprayoon, Pornpimon Palingoon, Kanokorn Trakultaweekoon, Supon Klaithin, Choochart Haruechaiyasak, Alisa Kongthon, Sumonmas Thatpitakul, and Sawit Kasuriya</i>	
Author Index	523