

Communications in Computer and Information Science

668

Commenced Publication in 2007

Founding and Former Series Editors:

Alfredo Cuzzocrea, Dominik Ślęzak, and Xiaokang Yang

Editorial Board

Simone Diniz Junqueira Barbosa

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),
Rio de Janeiro, Brazil*

Phoebe Chen

La Trobe University, Melbourne, Australia

Xiaoyong Du

Renmin University of China, Beijing, China

Joaquim Filipe

Polytechnic Institute of Setúbal, Setúbal, Portugal

Orhun Kara

TÜBİTAK BİLGEM and Middle East Technical University, Ankara, Turkey

Igor Kotenko

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

Ting Liu

Harbin Institute of Technology (HIT), Harbin, China

Krishna M. Sivalingam

Indian Institute of Technology Madras, Chennai, India

Takashi Washio

Osaka University, Osaka, Japan

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

Muyun Yang · Shujie Liu (Eds.)

Machine Translation

12th China Workshop, CWMT 2016
Urumqi, China, August 25–26, 2016
Revised Selected Papers

Editors

Muyun Yang
Harbin Institute of Technology
Harbin
China

Shujie Liu
Microsoft Research Asia
Beijing
China

ISSN 1865-0929

ISSN 1865-0937 (electronic)

Communications in Computer and Information Science

ISBN 978-981-10-3634-7

ISBN 978-981-10-3635-4 (eBook)

DOI 10.1007/978-981-10-3635-4

Library of Congress Control Number: 2016963154

© Springer Nature Singapore Pte Ltd. 2016

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.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Following the previous successful workshops in the series, the 12th China Workshop on Machine Translation (CWMT) was held during August 25–26, 2016, in Urumqi, China. This workshop provides an opportunity for researchers and practitioners to communicate and exchange ideas, and aims to improve the research of machine translation in China.

We were absolutely thrilled that 76 submissions were submitted to the conference. All of them were carefully reviewed in a double-blind manner and each paper was assigned to at least two independent reviewers. Finally, 15 Chinese and nine English papers were accepted, yielding an overall acceptance rate of 31.6%. With an English version of one Chinese paper, this proceedings volume comprises ten publications. As is traditionally the case with CWMT, the papers cover a wide range of subjects, including statistical MT, hybrid MT, MT evaluation, post editing, alignment, as well as inducing bilingual knowledge from corpora.

This year, CWMT 2016 featured three keynote speeches delivered by renowned experts in the field of MT and four invited talks exploring edge-cutting technologies by young researchers. A common topic that was highlighted in these talks is the emergence of neural-based MT paradigms, seemingly overtaking the statistical MT approach that first appeared in this conference 11 years ago.

There are a number of people we would like to thank. Firstly, this conference would not have been possible without the enormous efforts of Dr. Chengqing Zong, Dr. Le Sun, Prof. Tiejun Zhao, Prof. Jingbo Zhu, and Prof. Xiaodong Shi. We would especially like to thank the Springer for publishing the proceedings again (for the second time in CWMT series). Secondly, our heartfelt thanks go to Prof. Zhang Min, the general chair, and the members of the Program Committee, who, as usual, did sterling work over and above what might have reasonably been expected from them. Finally, we would like to thank Dr. Yating Yang and her team in Xinjiang Technical Institute of Physics and Chemistry, CAS, whose excellent conference organization impressed the attendees of CWMT 2016.

August 2016

Muyun Yang

Organization

Steering Committee

Chengqing Zong	Institute of Automation of Chinese Academy of Sciences, China
Le Sun	Institute of Software Chinese Academy of Sciences, China
Tiejun Zhao	Harbin Institute of Technology, China
Jingbo Zhu	Northeastern University, China
Xiaodong Shi	Xiamen University, China

Conference Chair

Min Zhang	Soochow University, China
-----------	---------------------------

Program Chair

Muyun Yang	Harbin Institute of Technology, China
------------	---------------------------------------

Publication Chair

Shujie Liu	Microsoft Research Asia
------------	-------------------------

Program Committee

Hailong Cao	Harbin Institute of Technology, China
Yidong Chen	Xiamen University, China
Yufeng Chen	Beijing Jiaotong University, China
Chong Feng	Beijing Institute of Technology, China
Yuhang Guo	Beijing Institute of Technology, China
Yanqing He	Institute of Scientific and Technical Information of China, China
Zhongjun He	Baidu, Inc.
Shujian Huang	Nanjing University, China
Wenbin Jiang	Institute of Computing Technology, Chinese Academy of Sciences, China
Hongfei Jiang	Dinfo Beijing Technology Development Co., Ltd., China
Jianfeng Li	USTC iflytek Co., Ltd.
Lemao Liu	National Institute of Information and Communications Technology, China
Shujie Liu	Microsoft Research Asia
Yang Liu	Tsinghua University
Weihua Luo	Alibaba Inc.
Cunli Mao	Kunming University of Science and Technology

Haitao Mi	IBM T.J. Watson Research Center, USA
Hideya Mino	National Institute of Information and Communications Technology
Jinsong Su	Xiamen University, China
Zhaopeng Tu	Noah's Ark Lab
Tong Xiao	Northeastern University
Deyi Xiong	Soochow University, China
Mo Yu	IBM T.J. Watson Research Center, USA
Yating Yang	Xinjiang Technical Institute of Physics and Chemistry Chinese Academy of Sciences, China
Conghui Zhu	Harbin Institute of Technology, China
Hao Zhang	Google, Inc.
Jiajun Zhang	Institute of Automation of Chinese Academy of Sciences, China
Yu Zhou	Institute of Automation of Chinese Academy of Sciences, China
Yun Zhu	Beijing Normal University, China

Local Organization Chair

Yating Yang	Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, China
-------------	--

Local Organizing Committee

Xiaobo Wang	Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, China
Wei Yang	Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, China
Rui Dong	Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, China
Chenggang Mi	Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, China
Kamali	Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, China

Organizers

Chinese Information Processing Society of China



Co-organizer

Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences



Sponsors



GTC Technology Co., Ltd.



Shenyang YaTrans Network Technology Co., Ltd.



Guangxi Daring E-Commerce Services Co., Ltd.



Beijing Lingsail Tech Co., Ltd.

Contents

MinKSR: A Novel MT Evaluation Metric for Coordinating Human Translators with the CAT-Oriented Input Method	1
<i>Guoping Huang, Chunlu Zhao, Hongyuan Ma, Yu Zhou, and Jiajun Zhang</i>	
Pivot-Based Semantic Splicing for Neural Machine Translation.	14
<i>Di Liu, Conghui Zhu, Tiejun Zhao, Xiaoxue Wang, and Muyun Yang</i>	
Re-ranking for Bilingual Lexicon Extraction with Bi-directional Linear Transformation from Comparable Corpora	25
<i>Chunyue Zhang and Tiejun Zhao</i>	
Learning Bilingual Sentence Representations for Quality Estimation of Machine Translation	35
<i>Junguo Zhu, Muyun Yang, Sheng Li, and Tiejun Zhao</i>	
Research on Domain Adaptation for SMT Based on Specific Domain Knowledge.	43
<i>Yanqing He, Liang Ding, and Ying Li</i>	
Automatic Construction of Domain Terminology Knowledge Base for HowNet Based on the Headword	61
<i>Chuang Wu, Lin Wang, Na Ye, Guiping Zhang, and Dongfeng Cai</i>	
Building the Vietnamese Phrase Treebank by Improved Probabilistic Context-Free Grammars	75
<i>Ying Li, Jianyi Guo, Zhengtao Yu, Yantuan Xian, and Yonghua Wen</i>	
Classifying Commas for Patent Machine Translation	91
<i>Hongzheng Li and Yun Zhu</i>	
BLEUS-syn: Cilin-Based Smoothed BLEU.	102
<i>Junting Yu, Wuying Liu, Hongye He, and Mianzhu Yi</i>	
Research on the Calculation Method of Semantic Similarity Based on Concept Hierarchy	113
<i>Kai Wang</i>	
Author Index	125