

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

Alfredo Cuzzocrea

ICAR-CNR and University of Calabria, Italy

Xiaoyong Du

Renmin University of China, Beijing, China

Joaquim Filipe

Polytechnic Institute of Setúbal, Portugal

Orhun Kara

TÜBİTAK BİLGEM and Middle East Technical University, Turkey

Igor Kotenko

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

Krishna M. Sivalingam

Indian Institute of Technology Madras, India

Dominik Ślęzak

University of Warsaw and Infobright, Poland

Takashi Washio

Osaka University, Japan

Xiaokang Yang

Shanghai Jiao Tong University, China

Xiaodong Shi Yidong Chen (Eds.)

Machine Translation

10th China Workshop, CWMT 2014
Macau, China, November 4-6, 2014
Proceedings

Volume Editors

Xiaodong Shi
Xiamen University, Fujian, China
E-mail: mandel@xmu.edu.cn

Yidong Chen
Xiamen University, Fujian, China
E-mail: ydchen@xmu.edu.cn

ISSN 1865-0929

e-ISSN 1865-0937

ISBN 978-3-662-45700-9

e-ISBN 978-3-662-45701-6

DOI 10.1007/978-3-662-45701-6

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014955083

© Springer-Verlag Berlin Heidelberg 2014

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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The 10th China Workshop on Machine Translation (CWMT), which was held during November 4–6, 2014, in Macau, aimed to bring together researchers and practitioners in the field of machine translation (MT) and provide a forum to facilitate communication and exchange among domestic and foreign scholars on the latest developments in the field of MT, with a special emphasis on the languages in China.

Following the success of the first CWMT in Xiamen in 2005, nine successful workshops were organized, including five Machine Translation evaluations (2007, 2008, 2009, 2011, 2013), an open source systems' development (2006), and two strategic meetings (2010, 2012). These activities have made a substantial impact on fostering the research and development of MT technologies in China. CWMT is currently considered as a leading and important academic event in the natural language processing field in China.

The first Machine Translation Evaluation and Workshop was held successfully in Xiamen University, jointly organized by: the Institute of Automation, Chinese Academy of Sciences; Institute of Computing Technology, Chinese Academy of Sciences; and Xiamen University. Subsequent workshops were jointly organized by the Institute of Computing Technology, Chinese Academy of Sciences; Institute of Automation, Chinese Academy of Sciences; Institute of Software, Chinese Academy of Sciences; Harbin Institute of Technology; and Xiamen University. CWMT was organized successfully in different places between 2006 and 2013, including: Xiamen University; the Institute of Computing Technology, Chinese Academy of Sciences; Harbin Institute of Technology; Institute of Automation, Chinese Academy of Sciences; Nanjing University; Institute of Software, Chinese Academy of Sciences; Xi'an University of Technology; and Kunming University of Science and Technology. This year, CWMT 2014 was organized by Macau University.

CWMT 2014 featured keynote speeches delivered by renowned experts in the field of MT; panel discussions and tutorials exploring leading-edge technologies and applications in MT; and a best student paper award given during the conference. An exhibition of commercial and research systems was held during the conference bringing together the users, developers, and researchers of MT. This conference offered them a great opportunity to share and exchange valuable experiences in MT, and thus to advance MT research and development in the region.

This year's paper submission was made through the EasyChair conference system. We received a total of 29 papers, among which 27 were from China. All the papers are carefully reviewed in a double-blind manner and each paper was assigned to three independent reviewers.

It was the first time that CWMT used EasyChair for the review phase. The review process is divided into four steps:

1. After the submission deadline has passed, all the reviewers are notified so that they can bid for the papers they like to review most (5 days). Most of the reviewers are Program Committee members of the previous CWMT workshops.
2. The reviewing process begins (15 days). Every paper is assigned to three reviewers. If a paper is not chosen by any reviewer, it is assigned randomly by EasyChair.
3. The first round of the review process ends, and the author response period (5 days) begins.
4. The reviews may change the accept/reject decision according to the author response. And in some cases, the Program Committee chair is involved. After 7 days, authors are then notified about the status of their papers.

Finally, 11 Chinese and 11 English papers were accepted (one accepted English paper was withdrawn at the end). The accepted papers address all aspects of MT, including preprocessing, language modeling, MT models, MT evaluation, and application to the patent translation, with a focus on MT of Chinese from/to minority languages. We believe that the publication and presentations of all these papers will advance the state of the art of MT research in China and on Chinese.

Last but not least, this is the first time that the conference proceedings (English manuscripts) of CWMT are published by Springer in the *Communications in Computer and Information Science* series, which we hope will draw more attention to this conference.

October 2014

Xiaodong Shi

Organization

Honorary Chair

Philip C.L. Chen University of Macau

Conference Chair

Tiejun Zhao Harbin Institute of Technology

Program Chair

Xiaodong Shi Xiamen University

Tutorial Chair

Qun Liu Institute of Computing Technology, Chinese
Academy of Sciences (CAS-ICT)

Panel Discussion Chair

Jinbo Zhu Northeastern University

Publication Chair

Yidong Chen Xiamen University

Program Committee (sorted according to Pinyin)

Dongfeng Cai	Shenyang Institute of Aeronautical Engineering
Cai Rangjia	Qinghai Normal University
Hailong Cao	Harbin Institute of Technology
Wenhan Chao	Beihang University
Boxing Chen	ICT-NRC, Canada
Jiajun Chen	Nanjing University
Yidong Chen	Xiamen University
Jinhua Du	Xi'an University of Technology
Xiangyu Duan	Suzhou University
Guohong Fu	Heilongjiang University
Yanqing He	Institute of Scientific and Technical Information of China
Zhongjun He	Baidu

VIII Organization

Degen Huang	Dalian University of Technology
Yeyan Huang	Beijing Institute of Technology
Hongfei Jiang	National Patent Information Center
Wenbin Jiang	Institute of Computing Technology, Chinese Academy of Sciences (CAS-ICT)
Mu Li	Microsoft Research Asia
Qun Liu	Institute of Computing Technology, Chinese Academy of Sciences (CAS-ICT)
Yang Liu	Tsinghua University
YanJun Ma	Baidu
Yao Meng	Fujitsu Research & Development Center
Na Shunwuritu	Mongolia University
Xiaodong Shi	Xiamen University
Le Sun	Institute of Software, Chinese Academy of Sciences (CAS-IS)
Türgün Ibrahim	Xinjiang University
Haifeng Wang	Baidu
Hua Wu	Baidu
Fei Xia	University of Washington
Tong Xiao	Northeastern University
Deyi Xiong	Suzhou University
Jinan Xu	Beijing Jiaotong University
Endong Xun	Beijing Language and Culture University
Muyun Yang	Harbin Institute of Technology
Yusup Abaydul	Xinjiang Normal University
Zhengtao Yu	Kunming University
Jiajun Zhang	Institute of Automation, Chinese Academy of Sciences (CAS-IA)
Min Zhang	Soochow University
Yujie Zhang	Beijing Jiaotong University
Hongmei Zhao	Institute of Computing Technology, Chinese Academy of Sciences (CAS-ICT)
Tiejun Zhao	Harbin Institute of Technology
Yu Zhou	Institute of Automation, Chinese Academy of Sciences (CAS-IA)
Jingbo Zhu	Northeastern University
Chengqing Zong	Institute of Automation, Chinese Academy of Sciences

Local Organization Chairs

Derek F. Wong	University of Macau
Lidia S. Chao	University of Macau

Local Organizing Committee

Francisco Oliveira	University of Macau
Hoklam Sio	University of Macau
Wenghong Lam	University of Macau
Xiaodong Zeng	University of Macau
Liang Tian	University of Macau
Longye Wang	University of Macau

Organizers

Chinese Information Processing Society of China



Co-organizers

University of Macau



Faculty of Science and Technology, University of Macau



Sponsors

Fundo para o Desenvolvimento das Ciências e da Tecnologia



科學技術發展基金
F | D | C | T

Table of Contents

Making Language Model as Small as Possible in Statistical Machine Translation	1
<i>Yang Liu, Jiajun Zhang, Jie Hao, and Dakun Zhang</i>	
Data Selection via Semi-supervised Recursive Autoencoders for SMT Domain Adaptation	13
<i>Yi Lu, Derek F. Wong, Lidia S. Chao, and Longyue Wang</i>	
Effective Hypotheses Re-ranking Model in Statistical Machine Translation	24
<i>Yiming Wang, Longyue Wang, Derek F. Wong, and Lidia S. Chao</i>	
Recognizing and Reordering the Translation Units in a Long NP for Chinese-English Patent Machine Translation	33
<i>Xiaodie Liu, Yun Zhu, and Yaohong Jin</i>	
A Statistical Method for Translating Chinese into Under-resourced Minority Languages	49
<i>Lei Chen, Miao Li, Jian Zhang, Zede Zhu, and Zhenxin Yang</i>	
Character Tagging-Based Word Segmentation for Uyghur	61
<i>Yating Yang, Chenggang Mi, Bo Ma, Rui Dong, Lei Wang, and Xiao Li</i>	
Analysis of the Chinese - Portuguese Machine Translation of Chinese Localizers <i>Qian</i> and <i>Hou</i>	70
<i>Chunhui Lu, Ana Leal, Paulo Quaresma, and Márcia Schmaltz</i>	
Chunk-Based Dependency-to-String Model with Japanese Case Frame	80
<i>Jinan Xu, Peihao Wu, Jun Xie, and Yujie Zhang</i>	
A Novel Hybrid Approach to Arabic Named Entity Recognition	93
<i>Mohamed A. Meselhi, Hitham M. Abo Bakr, Ibrahim Ziedan, and Khaled Shaalan</i>	
Reexaminatin on Voting for Crowd Sourcing MT Evaluation	104
<i>Yiming Wang and Muyun Yang</i>	
Author Index	117