Lecture Notes in Computer Science 3869

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Steve Renals Samy Bengio (Eds.)

Machine Learning for Multimodal Interaction

Second International Workshop, MLMI 2005
Edinburgh, UK, July 11-13, 2005
Revised Selected Papers

Volume Editors

Steve Renals
University of Edinburgh, Centre for Speech Technology Research
2 Buccleuch Place, Edinburgh EH8 9LW, UK
E-mail: s.renals@ed.ac.uk

Samy Bengio
IDIAP Research Institute
Rue du Simplon 4, Case Postale 592, 1920 Martigny, Switzerland
E-mail: bengio@idiap.ch

Library of Congress Control Number: 2006920577

CR Subject Classification (1998): H.5.2-3, H.5, I.2.6, I.2.10, I.2, I.7, K.4, I.4

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-540-32549-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-32549-9 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11677482 06/3142 5 4 3 2 1 0

Preface

This book contains a selection of refereed papers presented at the Second Workshop on Machine Learning for Multimodal Interaction (MLMI 2005), held in Edinburgh, Scotland, during 11–13 July 2005.

The workshop was organized and sponsored jointly by two European integrated projects, three European Networks of Excellence and a Swiss national research network:

- AMI, Augmented Multiparty Interaction, http://www.amiproject.org/
- CHIL, Computers in the Human Interaction Loop, http://chil.server.de/
- HUMAINE, Human–Machine Interaction Network on Emotion, http://emotion-research.net/
- PASCAL, Pattern Analysis, Statistical Modeling and Computational Learning, http://www.pascal-network.org/
- SIMILAR, human–machine interfaces similar to human–human communication, http://www.similar.cc/
- IM2, Interactive Multimodal Information Management, http://www.im2.ch/

In addition to the main workshop, MLMI 2005 hosted the NIST (US National Institute of Standards and Technology) Meeting Recognition Workshop. This workshop (the third such sponsored by NIST) was centerd on the Rich Transcription 2005 Spring Meeting Recognition (RT-05) evaluation of speech technologies within the meeting domain. Building on the success of the RT-04 spring evaluation, the RT-05 evaluation continued the speech-to-text and speaker diarization evaluation tasks and added two new evaluation tasks: speech activity detection and source localization.

MLMI 2005 was thus sponsored by the European Commission (Information Society Technologies priority of the Sixth Framework Programme), the Swiss National Science Foundation and the US National Institute of Standards and Technology.

Given the multiple links between the above projects and several related research areas, and the success of the first MLMI 2004 workshop, it was decided to organize once again a joint workshop bringing together researchers from the different communities working around the common theme of advanced machine learning algorithms for processing and structuring multimodal human interaction. The motivation for creating such a forum, which could be perceived as a number of papers from different research disciplines, evolved from an actual need that arose from these projects and the strong motivation of their partners for such a multidisciplinary workshop. This assessment was confirmed this year by a significant increase in the number of sponsoring research projects, and by the success of the workshop itself, which attracted about 170 participants.

The conference program featured invited talks, full papers (subject to careful peer review, by at least three reviewers), and posters (accepted on the basis of

abstracts) covering a wide range of areas related to machine learning applied to multimodal interaction — and more specifically to multimodal meeting processing, as addressed by the various sponsoring projects. These areas included:

- Human–human communication modeling
- Speech and visual processing
- Multimodal processing, fusion and fission
- Multimodal dialog modeling
- Human–human interaction modeling
- Multimodal data structuring and presentation
- Multimedia indexing and retrieval
- Meeting structure analysis
- Meeting summarizing
- Multimodal meeting annotation
- Machine learning applied to the above

Out of the submitted full papers, about 50% were accepted for publication in the present volume, after having been invited to take review comments and conference feedback into account.

In the present book, and following the structure of the workshop, the papers are divided into the following sections:

1. Invited Papers
2. Multimodal Processing
3. HCI and Applications
4. Discourse and Dialog
5. Emotion
6. Visual Processing
7. Speech and Audio Processing
8. NIST Meeting Recognition Evaluation

Based on the successes of MLMI 2004 and MLMI 2005, it was decided to organize MLMI 2006 in the USA, in collaboration with NIST (US National Institute of Standards and Technology), again in conjunction with the NIST meeting recognition evaluation.

Finally, we take this opportunity to thank our Program Committee members, the sponsoring projects and funding agencies, and those responsible for the excellent management and organization of the workshop and the follow-up details resulting in the present book.

November 2005

Steve Renals
Samy Bengio

Organization

General Chairs

Steve Renals	University of Edinburgh
Samy Bengio	IDIAP Research Institute

Local Organization

Caroline Hastings	University of Edinburgh
Avril Heron	University of Edinburgh
Bartosz Dobrzelecki	University of Edinburgh
Jean Carletta	University of Edinburgh
Mike Lincoln	University of Edinburgh

Program Committee

Marc Al-Hames	Munich University of Technology
Tilman Becker	DFKI
Hervé Bourlard	IDIAP Research Institute
Jean Carletta	University of Edinburgh
Franciska de Jong	University of Twente
John Garofolo	NIST
Thomas Hain	University of Sheffield
Lori Lamel	LIMSI
Benoit Macq	UCL-TELE
Johanna Moore	University of Edinburgh
Laurence Nigay	CLIPS-IMAG
Barbara Peskin	ICSI
Thierry Pun	University of Geneva
Marc Schröder	DFKI
Rainer Stiefelhagen	Universitaet Karlsruhe

NIST Meeting Recognition Workshop Organization

Jon Fiscus	NIST
John Garofolo	NIST

Sponsoring Projects and Institutions

Projects:

- AMI, Augmented Multiparty Interaction, http://www.amiproject.org/
- CHIL, Computers in the Human Interaction Loop, http://chil.server.de/
- HUMAINE, Human–Machine Interaction Network on Emotion, http://emotion-research.net/
- SIMILAR, human–machine interfaces similar to human–human communication, http://www.similar.cc/
- PASCAL, Pattern Analysis, Statistical Modeling and Computational Learning, http://www.pascal-network.org/
- IM2, Interactive Multimodal Information Management, http://www.im2.ch/

Institutions :

- European Commission, through the Multimodal Interfaces objective of the Information Society Technologies (IST) priority of the Sixth Framework Programme.
- Swiss National Science Foundation, through the National Center of Competence in Research (NCCR) program.
- US National Institute of Standards and Technology (NIST), http://www.nist.gov/speech/

Table of Contents

III HCI and Applications

IV Discourse and Dialogue

V Emotion

VI Visual Processing

VII Speech and Audio Processing

VIII NIST Meeting Recognition Evaluation