Lecture Notes in Electrical Engineering

Volume 427

About this Series

"Lecture Notes in Electrical Engineering (LNEE)" is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer's high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer's other Lecture Notes series, LNEE will be distributed through Springer's print and electronic publishing channels.

More information about this series at http://www.springer.com/series/7818

Kristiina Jokinen · Graham Wilcock
Editors

Dialogues with Social Robots

Enablements, Analyses, and Evaluation

Springer

Editors
Kristiina Jokinen
University of Helsinki
Helsinki
Finland

Graham Wilcock
University of Helsinki
Helsinki
Finland

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-10-2584-6 ISBN 978-981-10-2585-3 (eBook)
DOI 10.1007/978-981-10-2585-3

Library of Congress Control Number: 2016952821

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, #22-06/08 Gateway East, Singapore 189721, Singapore

The original version of this book was revised: Book volume number has been changed from 999 to 427. The Erratum to this chapter is available at 10.1007/978-981-10-2585-3_40

Preface

Human–robot interaction (HRI) is a growing area in several research communities, and its research topics cut across various fields in interaction technology, speech technology, computer science, cognitive science, neuroscience, engineering, and many other related disciplines. It is expected that robots will play a bigger role in future environments and societies, and in particular their capabilities for natural language communication will increase: besides being able to perform various physical tasks, the robot will understand human social signals and support extensive, flexible, adaptable, and reliable spoken language interaction. Such communicating robot agents are generally called *social robots*, and they will provide interactive interfaces to digital data, adapting to the needs and requirements of different users.

In spoken dialogue research, interaction with robots is an increasingly active research and development area in both academia and industry. This is understandable, because intelligent communicative agents have been an application area for dialogue management since the late 1970s, when the first Belief–Desire–Intention (BDI) agents were developed with plan-based dialogue modelling as the mainstream technique. Nowadays spoken dialogue technology is ripe enough to be integrated in the robot platforms, while statistical and rule-based dialogue modelling techniques allow versatile experimentation with natural multimodal communication strategies between humans and intelligent agents. On the other hand, robot applications also provide useful testing grounds for applying dialogue models in practical contexts, as well as evaluating usability of spoken dialogue systems in real contexts. It should be noticed that "intelligent agent" here refers both to robots (situated agents able to move in the physical environment) and to animated and virtual agents (embodied conversational agents that appear on the screen).

Although the title of the book refers to robots, the research area also concerns chat systems. These are popular text-based dialogue systems whose development draws on dialogue modelling, and which can provide tools for data collection and further studies on dialogue management.

As an emergent field of research, human–robot interaction has many open questions. Important topics concern various aspects in multimodal communication,

open-domain dialogue management, emotion detection, social acceptability, and natural interaction modelling. One of the major questions is how to evaluate and benchmark human–robot dialogues, and how to best model issues related to a social robot's dialogue capabilities. Studies need data from natural interaction situations, and the data is needed for two purposes: to build appropriate models for dialogue interactions, and to evaluate the models and modules, and the interaction as a whole with an intelligent agent system.

Communication via novel interfaces, especially with situated social robots, does not only involve interaction with an information-providing tool but also includes aspects which are typical of human–human communication: timing of responses, feedback-giving strategies, error management, use of multimodal behaviour, and understanding of the partner's contributions. The effect of a human-like appearance and the acceptability of the application in general are also crucial topics brought in by the development of novel autonomous interactive systems. It is important to study how dialogue strategies are carried over into new situations and new applications, i.e. to explore *affordance* of social robots. Moreover, advances in research and development of autonomous situated agents propel discussions on the ethics of social robots: their decision-making, application domains, conflicting goals between robots and their users, and the responsibilities of researchers. Such issues elevate research and development on social robotics to another level of social research: that of improving human well-being.

The book describes work on spoken dialogue systems and intelligent social agents. The chapters explore novel ways to model and use interaction strategies, and present improvements and new developments beyond previous work. The contributions are selected, adapted, and expanded from the papers presented at the Seventh International Workshop on Spoken Dialogue Systems (IWSDS 2016), and they offer different perspectives and solutions for the important questions, through papers by leading researchers covering key topics in the field. We hope the book will contribute to the design and study of interaction patterns that support successful human–robot communication, and will encourage ethical discussion and evaluation of the applications.

The IWSDS 2016 Conference

The series of IWSDS conferences aims to bring together researchers from both industry and academia working within the various disciplines that constitute the field of spoken dialogue systems. The Seventh IWSDS (IWSDS 2016) was held from January 13 to 16, 2016 at Hotel Riekonlinna in Saariselkä, Finland. This was the northernmost spoken dialogue workshop ever, and it allowed participants to discuss implementation and analytical work as well as to enjoy beautiful winter scenes and activities in Finnish Lapland. The *Kaamos*, the Arctic night when the sun stays below the horizon, had just ended, and at the time of the conference the daylight was less than three hours long; however, it contributed to wonderful sunrises and sunsets.

The special theme of IWSDS 2016 was *Dialogues with Social Robots: Enablements, Analyses, and Evaluation*. The full programme included presentations and demonstrations, a gala dinner and a trip to Inari to learn about the indigenous Sami culture at the Siida Museum for the Sami people and at Sajos, the Sami Parliament in Finland.

The conference also featured two invited speakers. Riitta Hari (Aalto University) spoke *About the Primacy of Social Interaction in Human Brain Function*, describing how current neuroscience, even "social neuroscience", examines brain functions of isolated persons who observe other persons' actions from a third-person view. Professor Hari advocated the idea that research on the brain basis of social cognition and interaction should move from this "spectator science" to studies of engaged participants. Recent advances in neuroimaging now allow study of the brain basis of *social interaction* by simultaneous recordings of brain activity of two participants engaged in natural communication. An important research question is whether social interaction emerges from lower-level perceptual, motor and cognitive functions, as is typically assumed, or whether social interaction is in fact the primary default mode governing perception and action, which would challenge many current ideas about human brain function.

David Traum (ICT, University of Southern California) spoke on *The Role of a Lifetime: Dialogue Models for Virtual Human Role-players*. Some dialogue systems are instruments to allow a user to efficiently solve a task, but other dialogue systems act as role-players in learning exercises, games, or other activities. For these cases, human-like dialogue is more important, especially if the purpose of the dialogue activity is to learn or practice interactive skills that should transfer to other humans. However, for role-play dialogue the most appropriate metric is not the familiar "Turing test" of indistinguishability from humans, but rather activation of the same sociocognitive skills that are appropriate for human interaction. Professor Traum presented examples of role-play dialogue systems from a wide variety of activities, genres, and roles, focussing on virtual humans created at the USC Institute for Creative Technologies.

Besides the normal sessions of papers and posters, the conference included three special sessions: *Dialogue State Tracking Challenge 4*, *Evaluation of Human–Robot Dialogue in Social Robotics*, and *Sociocognitive Language Processing*, which are briefly described below.

Dialogue State Tracking Challenge 4

Organisers: Luis F. D'Haro, Seokhwan Kim, Rafael E. Banchs, Matthew Henderson and Jason Williams.

Dialogue state tracking is one of the key sub-tasks of dialogue management, which defines the representation of dialogue states and updates them at each moment in a given ongoing conversation. In this challenge, participants used the TourSG corpus to develop the components. TourSG consists of dialogue sessions

about touristic information for Singapore collected from Skype calls between 3 tour guides and 35 tourists. The challenge included a main task on dialogue state tracking at sub-dialogue level; four optional pilot tasks on spoken language understanding, speech act prediction, spoken language generation, and the implementation of end-to-end systems; and an optional open track for participants to explore any task of their interest over the provided dataset.

Evaluation of Human–Robot Dialogue in Social Robotics

Organisers: Laurence Devillers, Kristiina Jokinen, Joseph Mariani, Haizhu Li, Alex Waibel and Wolfgang Minker.

The aim of this special theme was to make a bridge between the social robotics and spoken dialogue research communities. In the spoken dialogue community the focus is on evaluation of verbal interaction including spontaneous speech recognition and understanding, whereas in the social robotics community the focus is on evaluation of engagement measures in interaction, with non-verbal features such as acoustics and gestures but without spontaneous speech information. Several robotics competitions already incorporate human-robot interaction (such as RoboCup@Home, which has benchmarks to evaluate robots in realistic home environments, or the AAAI Grand Challenge in which the robot attends a conference and delivers a talk) but these competitions do not focus on interactive spoken dialogues. Similarly, several spoken dialogue competitions have been proposed by Darpa (such as ATIS or Communicator), but as yet there are no established protocols and no ongoing evaluation campaigns for human-robot interaction.

Sociocognitive Language Processing

Organisers: Björn Schuller and Michael McTear.

Sociocognitive Language Processing (SCLP) is the idea of coping with everyday language, including slang and multilingual phrases and cultural aspects, and in particular with irony/sarcasm/humour and paralinguistic information such as the physical and mental state and traits of the dialogue partner (affect, age group, personality dimensions) and social aspects. Multimodal aspects such as facial expression, gestures, or bodily behaviour should also be included in the analysis. At the same time, SCLP can render future dialogue systems more “chatty” by not only feeling natural but also being emotionally and socially competent, ideally leading to a more symmetrical dialogue. For that, the computer should itself have a “need for humour” and an “increase of familiarity”, enabling computers to experience or at least better understand emotions and personality.

We warmly thank the invited speakers, the steering committee, the reviewers in the scientific committee, the organisers of the special sessions, the members of the local arrangements committee, and all the participants for their contributions to the success of IWSDS 2016.

Helsinki, Finland
July 2016

Kristiina Jokinen
Conference Chair IWSDS 2016
Graham Wilcock
Local Arrangements Chair IWSDS 2016

Organization of IWSDS 2016

Steering Committee

Gary Geunbae Lee, POSTECH, Pohang, Korea
Ramón López-Cózar, University of Granada, Spain
Joseph Mariani, LIMSI and IMMI-CNRS, Orsay, France
Wolfgang Minker, Ulm University, Germany
Satoshi Nakamura, NAIST, Nara, Japan

Conference Chair

Kristiina Jokinen, University of Helsinki, Finland, and University of Tartu, Estonia

Local Arrangements Committee

Katri Hiovain, University of Helsinki, Finland
Mikko Kurimo, Aalto University, Finland
Niklas Laxström, University of Helsinki, Finland
Juho Leinonen, Aalto University, Finland
Ilona Rauhala, University of Helsinki, Finland
Markku Turunen, University of Tampere, Finland
Graham Wilcock, University of Helsinki, Finland

Scientific Committee

Samer Al Moubayed, Disney Research, USA
Masahiro Araki, Kyoto Institute of Technology, Japan
Rafael E. Banchs, Institute for Infocomm Research, Singapore
Frédéric Béchet, Aix-Marseille University, France
Felix Burkhard, Deutsche Telekom, Germany
Zoraida Callejas, University Granada, Spain
Nick Campbell, Trinity College Dublin, Ireland
Heriberto Cuayahuitl, Heriot-Watt University, UK
Ramón López-Cózar, University of Granada, Spain
Laurence Devillers, LIMSI-CNRS, France
Luis Fernando D'Haro, Institute for Infocomm Research, Singapore
Giuseppe Di Fabbrizio, Amazon Research, USA
Jens Edlund, KTH, Sweden
Julien Epps, The University of New South Wales, Australia
Benoit Favre, Aix-Marseille University, France
Milica Gašić, Cambridge University, UK
Kallirroi Georgila, ICT USC, USA
David Griol, Universidad Carlos III de Madrid, Spain
Dilek Hakkani-Tur, Microsoft Research, USA
Matthew Henderson, Google, USA
Ryuichiro Higashinaka, NTT, Japan
Anna Hjalmarsson, KTH, Sweden
Kristiina Jokinen, University of Helsinki, Finland
Pamela Jordan, University of Pittsburgh, USA
Tatsuya Kawahara, Kyoto University, Japan
Hong Kook Kim, Gwangju Institute of Science and Technology, Korea
Seokhwan Kim, Institute for Infocomm Research, Singapore
Kazunori Komatani, Osaka University, Japan
Stefan Kopp, Bielefeld University, Germany
Mikko Kurimo, Aalto University, Finland
Fabrice Lefèvre, University of Avignon, France
Rivka Levitan, Brooklyn College, USA
Haizhou Li, Institute for Infocomm Research, Singapore
Joseph Mariani, LIMSI & IMMI-CNRS, Orsay, France
Michael McTear, University of Ulster, UK
Wolfgang Minker, Ulm University, Germany
Teruhisa Misu, Honda Research Institute, USA
Satoshi Nakamura, Nara Institute of Science and Technology, Japan
Mikio Nakano, Honda Research Institute, Japan
Olivier Pietquin, University of Lille, France
Paul Piwek, The Open University, UK
Matthew Purver, Queen Mary University of London, UK

Sakriani Sakti, Nara Institute of Science and Technology, Japan
David Schlangen, Bielefeld University, Germany
Björn Schuller, University of Passau, Germany
Gabriel Skantze, KTH, Sweden
Amanda Stent, Yahoo! Labs, USA
Svetlana Stoyanchev, AT&T Labs Research, USA
David Suendermann-Oeft, ETS Research, USA
David Traum, ICT USC, USA
Markku Turunen, University of Tampere, Finland
Carl Vogel, Trinity College Dublin, Ireland
Nigel Ward, University of Texas at El Paso, USA
Graham Wilcock, University of Helsinki, Finland
Jason D. Williams, Microsoft Research, USA

Sponsors

We gratefully thank the Federation of Finnish Learned Societies and the DigiSami Project for their financial support for the conference. The workshop was endorsed by the Special Interest Group on Discourse and Dialogue (SIGDIAL), the Robotics Society of Finland, and the International Speech Communication Association (ISCA).

FEDERATION OF FINNISH LEARNED SOCIETIES
Delegation of the Finnish Academies of Science and Letters

DIGISAMI

Contents

Part I The Northernmost Spoken Dialogue Workshop

DigiSami and Digital Natives: Interaction Technology for the North Sami Language 3
Kristiina Jokinen, Katri Hiovain, Niklas Laxström, Ilona Rauhala and Graham Wilcock

Part II Methods and Techniques for Spoken Dialogue Systems

A Comparative Study of Text Preprocessing Techniques for Natural Language Call Routing 23
Roman Sergienko, Muhammad Shan and Alexander Schmitt

Compact and Interpretable Dialogue State Representation with Genetic Sparse Distributed Memory 39
Layla El Asri, Romain Laroche and Olivier Pietquin

Incremental Human-Machine Dialogue Simulation 53
Hatim Khouzaimi, Romain Laroche and Fabrice Lefèvre

Active Learning for Example-Based Dialog Systems 67
Takuya Hiraoka, Graham Neubig, Koichiro Yoshino, Tomoki Toda and Satoshi Nakamura

Question Selection Based on Expected Utility to Acquire Information Through Dialogue 79
Kazunori Komatani, Tsugumi Otsuka, Satoshi Sato and Mikio Nakano

Separating Representation, Reasoning, and Implementation for Interaction Management: Lessons from Automated Planning 93
Mary Ellen Foster and Ronald P.A. Petrick

***SimpleDS*: A Simple Deep Reinforcement Learning Dialogue System** 109
Heriberto Cuayáhuitl

Breakdown Detector for Chat-Oriented Dialogue 119
Tomo Horii, Hideaki Mori and Masahiro Araki

User Involvement in Collaborative Decision-Making Dialog Systems 129
Florian Nothdurft, Pascal Bercher, Gregor Behnke and Wolfgang Minker

Part III Socio-Cognitive Language Processing

Natural Language Dialog System Considering Speaker's Emotion Calculated from Acoustic Features 145
Takumi Takahashi, Kazuya Mera, Tang Ba Nhat, Yoshiaki Kurosawa and Toshiyuki Takezawa

Salient Cross-Lingual Acoustic and Prosodic Features for English and German Emotion Recognition 159
Maxim Sidorov, Christina Brester, Stefan Ultes and Alexander Schmitt

Entropy-Driven Dialog for Topic Classification: Detecting and Tackling Uncertainty 171
Manex Serras, Naiara Perez, María Inés Torres and Arantza del Pozo

Evaluation of Question-Answering System About Conversational Agent's Personality 183
Hiroaki Sugiyama, Toyomi Meguro and Ryuichiro Higashinaka

Fisher Kernels on Phase-Based Features for Speech Emotion Recognition 195
Jun Deng, Xinzhou Xu, Zixing Zhang, Sascha Frühholz, Didier Grandjean and Björn Schuller

Part IV Towards Multilingual, Multimodal, Open Domain Spoken Dialogue Systems

Internationalisation and Localisation of Spoken Dialogue Systems 207
Niklas Laxström, Graham Wilcock and Kristiina Jokinen

A Multi-lingual Evaluation of the vAssist Spoken Dialog System. Comparing Disco and RavenClaw 221
Javier Mikel Olaso, Pierrick Milhorat, Julia Himmelsbach, Jérôme Boudy, Gérard Chollet, Stephan Schlögl and María Inés Torres

Multimodal HALEF: An Open-Source Modular Web-Based Multimodal Dialog Framework 233
Zhou Yu, Vikram Ramanarayanan, Robert Mundkowsky, Patrick Lange, Alexei Ivanov, Alan W. Black and David Suendermann-Oeft

Data Collection and Synchronisation: Towards a Multiperspective Multimodal Dialogue System with Metacognitive Abilities 245
Fasih Haider, Saturnino Luz and Nick Campbell

HELPR: A Framework to Break the Barrier Across Domains in Spoken Dialog Systems 257
Ming Sun, Yun-Nung Chen and Alexander I. Rudnicky

Towards an Open-Domain Social Dialog System 271
Maria Schmidt, Jan Niehues and Alex Waibel

Part V Evaluation of Human–Robot Dialogue in Social Robotics

Is Spoken Language All-or-Nothing? Implications for Future Speech-Based Human-Machine Interaction 281
Roger K. Moore

Toward a Context-Based Approach to Assess Engagement in Human-Robot Social Interaction 293
Laurence Devillers and Guillaume Dubuisson Duplessis

Extrinsic Versus Intrinsic Evaluation of Natural Language Generation for Spoken Dialogue Systems and Social Robotics 303
Helen Hastie, Heriberto Cuayáhuitl, Nina Dethlefs, Simon Keizer and Xingkun Liu

Engagement in Dialogue with Social Robots 313
Loredana Cerrato and Nick Campbell

Lend a Hand to Service Robots: Overcoming System Limitations by Asking Humans 321
Felix Schüssel, Marcel Walch, Katja Rogers, Frank Honold and Michael Weber

Dialogue with Robots to Support Symbiotic Autonomy 331
Andrea Vanzo, Danilo Croce, Emanuele Bastianelli, Guglielmo Gemignani, Roberto Basili and Daniele Nardi

Towards SamiTalk: A Sami-Speaking Robot Linked to Sami Wikipedia 343
Graham Wilcock, Niklas Laxström, Juho Leinonen, Peter Smit, Mikko Kurimo and Kristiina Jokinen

Part VI Dialogue Quality Assessment

Utterance Selection Using Discourse Relation Filter for Chat-oriented Dialogue Systems 355
Atsushi Otsuka, Toru Hirano, Chiaki Miyazaki, Ryuichiro Higashinaka, Toshiro Makino and Yoshihiro Matsuo

Analysis of Temporal Features for Interaction Quality Estimation 367
Stefan Ultes, Alexander Schmitt and Wolfgang Minker

Recurrent Neural Network Interaction Quality Estimation 381
Louisa Pragst, Stefan Ultes and Wolfgang Minker

An Evaluation Method for System Response in Chat-Oriented Dialogue System 395
Hideaki Mori, Atsushi Yasuda and Masahiro Araki

The Negotiation Dialogue Game 403
Romain Laroche and Aude Genevay

Estimation of User's Willingness to Talk About the Topic: Analysis of Interviews Between Humans 411
Yuya Chiba and Akinori Ito

Recognising Conversational Speech: What an Incremental ASR Should Do for a Dialogue System and How to Get There 421
Timo Baumann, Casey Kennington, Julian Hough and David Schlangen

Part VII Dialogue State Tracking Challenge 4

The Fourth Dialog State Tracking Challenge 435
Seokhwan Kim, Luis Fernando D'Haro, Rafael E. Banchs, Jason D. Williams and Matthew Henderson

Convolutional Neural Networks for Multi-topic Dialog State Tracking 451
Hongjie Shi, Takashi Ushio, Mitsuru Endo, Katsuyoshi Yamagami and Noriaki Horii

The MSIIP System for Dialog State Tracking Challenge 4 465
Miao Li and Ji Wu

Robust Dialog State Tracking for Large Ontologies 475
Franck Dernoncourt, Ji Young Lee, Trung H. Bui and Hung H. Bui

Erratum to: Dialogues with Social Robots E1
Kristiina Jokinen and Graham Wilcock

Author Index 487