Semantic Interpretation of Multi-Modal Human-Behaviour Data

Making Sense of Events, Activities, Processes

Mehul Bhatt^{1,2} · Kristian Kersting³

© Springer-Verlag GmbH Deutschland 2017

Abstract This special issue presents interdisciplinary research—at the interface of artificial intelligence, cognitive science, and human-computer interaction—focussing on the semantic interpretation of human behaviour. The special issue constitutes an attempt to highlight and steer foundational methods research in artificial intelligence, in particular knowledge representation and reasoning, for the development of human-centred cognitive assistive technologies. Of specific interest and focus have been application outlets for basic research in knowledge representation and reasoning and computer vision for the cognitive, behavioural, and social sciences.

Keywords Deep semantics · Multimodality · Human behaviour studies · Visual perception · Spatial cognition · Artificial intelligence · Cognitive science

1 About the Special Issue

The special issue of the *German Journal of Artificial Intelligence* (KI) focusses on and emphasises general methods and tools for activity and event-based semantic interpretation of

multi-modal sensory data relevant to a range of application domains and problem contexts where interpreting human behaviour is central. The overall motivation and driving theme of the special issue pertains to artificial intelligence based methods and tools that may serve a foundational purpose toward the high-level semantic interpretation of large-scale, dynamic, multi-modal sensory data, or data streams. A crucial focus of the special issue has been on foundational methods supporting the development of human-centred technologies and cognitive interaction systems aimed at assistance and empowerment, e.g. in everyday life and activity, professional problem solving for creative and analytical decision-making, planning etc.

1.1 Multi-Modal Event and Activity Interpretation

The multi-modality that is alluded to in the context of this special issue stems from an inherent synergistic value in the integrated processing and interpretation of a range of data sources that are common in the context of cognitive interaction systems, computational cognition, and human-computer interaction scenarios.

Multi-modal data-sources that may be envisaged include, but are not limited to, one or more of the following:

- Visuo-spatial imagery:
 - Image, video, video and depth (RGB-D), point-clouds.
 - Geospatial satellite imagery, remote sensing data, crowd-sourced data, survey data.
- Movement and interaction data:

✉ Mehul Bhatt
bhatt@uni-bremen.de; mehul.bhatt@oru.se
Kristian Kersting
kristian.kersting@cs.tu-dortmund.de

¹ Human-Centred Cognitive Assistance Lab. (HCC),
University of Bremen, Bremen, Germany

² Centre for Applied Autonomous Sensor Systems (AASS),
Örebro University, Örebro, Sweden

³ Technical University of Dortmund (DE), Dortmund,
Germany

- Indoor or outdoor settings pertaining to motion of people/things at arbitrary spatial and temporal scales.
- Sensory-motor data about interaction of people with things (e.g., in activities of everyday living).
- Neurphysiological and other human behaviour data:
 - Eye-tracking and related human behaviour data.
 - FMRI, EEG data occurring in medical computing, brain-computer interfaces etc.

A key emphasis in multi-modality is on AI-based integrated “perceptual sensemaking” with mixed symbolic, qualitative, quantitative data aimed at empirical or evidence-based studies, high-level analytics, knowledge discovery etc. The applied goals or end-results may be driven by assisting humans in decision-making, innovation, planning, design, control, automation etc.

1.2 Foundational Methods

This special issue emphasises systematically formalised integrative artificial intelligence methods and tools, for instance combining reasoning and learning, that enable declarative modelling, reasoning and query answering, relational learning, embodied grounding and simulation, etc. Broadly, the role of declarative abstraction, knowledge representation and reasoning, and neural-symbolic learning and inference from multi-modal sensory data is highly welcome.

Key topics that the special issue/its open call solicited include:

- Scene, event, and activity interpretation.
- Event stream reasoning.
- Commonsense scene perception.
- Declarative spatial reasoning.
- Cognitive vision, semantic Q/A with video.
- Integration of relational logic and statistical learning.
- Deep visuo-spatial semantics.
- Reasoning about space, actions, and change.
- Relational learning and knowledge discovery.
- Visuo-spatial computing.
- Embodied visuo-auditory cognition.
- Computational models of narrative.
- Computational cognitive systems.
- Cognition and natural interaction.
- Assistive technologies.

In response to an open call for papers, contributions were peer-reviewed across a two revision rounds.

2 Content

This special issue is composed of:

1. technical contributions,
2. a research project, and
3. an interview report emanating from discussions as part of the CoDesign 2017 Roundtable [1] at the HCC Lab., University of Bremen (Germany).

Published contributions emanate from research labs in Australia, Germany, Switzerland, and United States. Contributing labs encompass departments of computer science, cognitive science, psychology, digital humanities, and engineering design (1–3):

1. Technical Contributions

- Declarative Reasoning about Space and Motion with Video
Jakob Suchan
- Automated Interpretation of Eye-Hand Coordination in Mobile Eye Tracking Recordings
Moritz Mussnug, Daniel Singer, Quentin Lohmeyer, Mirko Meboldt
- Automatic Detection of Visual Search for the Elderly using Eye and Head Tracking Data
Michael Dietz, Daniel Schork, Ionut Damian, Anika Steinert, Marten Haesner, Elisabeth André
- Assigning Group Activity Semantics to Multi-Device Mobile Sensor Data: An Explanation-Based Perspective
Seng Loke, Amin Abkenar

2. Research Project

- Red Hen Lab: Dataset and Tools for Multimodal Human Communication Research
Jungseock Joo, Francis Steen, Mark Turner

3. Interview

- Cognition, Interaction, Design: Discussions as Part of the CoDesign 2017 Roundtable
Mehul Bhatt, James Cutting, Daniel Levin, Clayton Lewis

Overall, all accepted contributions within the special issue directly deal human behaviour interpretation; topics range from semantic interpretation of activities from video, to other kinds of sensors involving eye-tracking, head

movement tracking, and other kinds of mobile sensors concerned with spatio-temporal and physiological data.

The contribution by [7] focusses on developing general declarative methods for the semantic interpretation of video and other kinds of visual-spatial data; as described by Suchan:

Semantic interpretation of dynamic visuo-spatial imagery is central to a broad range of applications where computational systems have to make sense of human interactions, such as scene and movie understanding, cognitive robotics, human behaviour analysis, or smart environments.

Towards providing systematic solutions for “making-sense” of dynamic visual data, the contribution of [7] presents:

a commonsense theory of space and motion for representing and reasoning about motion patterns in video data, to perform declarative (deep) semantic interpretation of visuo-spatial sensor data, e.g., coming from object tracking, eye tracking data, movement trajectories. The theory has been implemented within Constraint Logic Programming to support integration into large scale AI projects.

Closely linked is the work by [6], which focusses on the automatic interpretation of specific gestural instances involving eye-hand co-ordination in the context of mobile eye-tracking. To quote [6]:

this work aims to automatically detect cognitive demanding phases in mobile eye tracking recordings. The approach presented combines the user’s perception (gaze) and action (hand) to isolate demanding interactions based upon a multi-modal feature level fusion.

The next work in line by [3] addresses an aspect of visual perception, namely, the visual search behaviour in the context of elderly groups; their emphasis is on automatic detection of visual search derived on the basis an integration of eye and head tracking during experimental activities performed by subjects involving visual search for objects in a real-world setting. To quote [3]:

in order to collect the necessary sensor data for the recognition of visual search, we develop a completely mobile eye and head tracking device specifically tailored to the requirements of older adults.

The line of “semantic interpretation” of visuo-spatial imagery is extended further, where [5] propose to take a general view of sensory data interpretation by assigning group activity semantics to sensory data based on an approach

rooted in generating explanations (the context here is multi-device mobile sensor data). To quote [5]:

sensor data from disparate sources can be aggregated and inferences can be made about the user, the user’s physical activities as well as the physical activities of the group the user is part of....this paper proposes an explanation-based perspective on reasoning about multi-device sensor data, and describes a framework called GroupSense that prototypes this idea.

Where the above were research articles, the remaining two contributions by [2] and [4] are position statements aimed at emphasising interdisciplinary research questions at the interface of AI, cognition, design, and human-behaviour studies. Whereas [4] presents a project report on the RedHen Lab, [2] presents a report based on a discussion and interview with psychologists James Cutting (Cornell University) and Daniel Levin (Vanderbilt University), and HCI specialist Clayton Lewis (University of Colorado, Boulder) held as part of the CoDesign 2017 initiative.

As [4] state in their project report:

researchers in the fields of AI and Communication both study human communication, but despite the opportunities for collaboration, they rarely interact... this article introduces Red Hen Lab with some possibilities for collaboration, demonstrating the utility of a variety of machine learning and AI-based tools and methods to fundamental research questions in multimodal human communication.

Finally, the interview report of [2] aims to:

emphasise the interplay of minds (human behaviour), media (design artefacts), and (assistive) technology from the viewpoints of artificial intelligence, spatial cognition, and spatial design.

The discussions amongst Bhatt, Cutting, Levin, and Lewis within the interview report will hopefully clearly bring out open questions and possibilities for addressing the confluence of computational, cognitive, and psychological sciences towards the study and semantic interpretation of human behaviour, and its implications for the design and implementation of human-centred cognitive assistive technologies, and computational cognitive system.

3 Related Scientific Forums

- International Joint Conference on Artificial Intelligence (IJCAI) www.ijcai.org

- Association for the Advancement of Artificial Intelligence (AAAI) www.aaai.org
- European Conference on Artificial Intelligence (ECAI) www.eurai.org/activities/ECAI_conferences
- International Conference on Multimodal Communication: Developing New Theories and Methods (ICMC) <https://sites.google.com/a/case.edu/icmc2017/>
- The European Workshop on Imagery and Cognition (EWIC) <http://ewic2016.parisdescartes.fr>
- International Conference on Spatial Information Theory (COSIT) <http://cosit.info>
- International Conference on: Human-Computer Interaction (INTERACT) www.interact2017.org
- Computational Models of Narrative (CMN) <http://narrative.csail.mit.edu/>
- International Conference on Spatial Cognition (ICSC) www.icsc-rome.org
- ACM Symposium on Applied Perception (ACM SAP) <http://sap.acm.org>
- Journal of Spatial Cognition and Computation www.icsc-rome.org
- Cognitive Science (CogSci) / Conference, and Journal <http://www.cognitivesciencesociety.org/>
- Linguistics Vanguard: A Multimodal Journal for the Language Sciences <https://www.degruyter.com/view/j/lingvan>
- The Distributed Little Red Hen Lab <https://sites.google.com/site/distributedlittleredhen/>
- CoDesign 2017 – The Bremen Summer of Cognition and Design <http://hcc.uni-bremen.de/codesign2017/>
- International School on: Human-Centred Computing (HCC 2016) — Minds. Experiences. Technologies. <http://hcc.uni-bremen.de/school2016>

particularly thank Joachim Hertzberg for his advice and liaison in his role as managing editor throughout the development of the special issue. The interview report included in this special issue became possible because of the tremendous support and contributions of Daniel Levin, James Cutting, Clayton Lewis: thank you very much! Mehul Bhatt also acknowledges the logistical support available via the CoDesign 2017 initiative in this context. As part of the blind peer-review process, the support of external referees across the two revision rounds has been very valuable. We remain grateful to all contributing authors for their persistence and effort across the two review rounds that were undertaken for the preparation of this special issue.

References

1. Bhatt M (2017) CoDesign 2017—the Bremen summer of cognition and design/coDesign roundtable, University of Bremen. <http://hcc.uni-bremen.de/codesign2017/roundtable/>. Accessed April–Sep 2017
2. Bhatt M, Cutting J, Levin D, Lewis C (2017) Cognition, interaction, design: discussions as part of the coDesign 2017 roundtable. KI—Künstliche Intelligenz, 2017. ISSN 1610-1987
3. Dietz M, Schork D, Damian I, Steinert A, Haesner M, André E (2017) Automatic detection of visual search for the elderly using eye and head tracking data. KI—Künstliche Intelligenz, Aug 2017. ISSN 1610-1987. doi:[10.1007/s13218-017-0502-z](https://doi.org/10.1007/s13218-017-0502-z)
4. Joo J, Steen F, Turner M (2017) Red Hen Lab: dataset and tools for multimodal human communication research. KI—Künstliche Intelligenz, 2017. ISSN 1610-1987
5. Loke S, Abkenar AB (2017) Assigning group activity semantics to multi-device mobile sensor data: an explanation-based perspective. KI—Künstliche Intelligenz, 2017. ISSN 1610-1987
6. Mussnug M, Singer D, Lohmeyer Q, Meboldt M (2017) Automated interpretation of eye–hand coordination in mobile eye tracking recordings. KI—Künstliche Intelligenz, Aug 2017. ISSN 1610-1987. doi:[10.1007/s13218-017-0503-y](https://doi.org/10.1007/s13218-017-0503-y)
7. Suchan J (2017) Declarative reasoning about space and motion with video. KI—Künstliche Intelligenz, Aug 2017. ISSN 1610-1987. doi:[10.1007/s13218-017-0504-x](https://doi.org/10.1007/s13218-017-0504-x)

Acknowledgements The guest editors gratefully acknowledge the KI Journal editorial board for their support of this special issue; we