Description Logics in Multimedia Reasoning

Leslie F. Sikos

Description Logics in Multimedia Reasoning

Springer

Leslie F. Sikos
Centre for Knowledge & Interaction Technologies
Flinders University
Adelaide
Australia

ISBN 978-3-319-54065-8 ISBN 978-3-319-54066-5 (eBook)
DOI 10.1007/978-3-319-54066-5

Library of Congress Control Number: 2017934914

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The immense and constantly growing number of videos urges efficient automated processing mechanisms for multimedia contents, which is a real challenge due to the huge semantic gap between what computers can automatically interpret from audio and video signals and what humans can comprehend based on cognition, knowledge, and experience. Low-level features, which correspond to local and global characteristics of audio and video signals, and low-level feature aggregates and statistics, such as various histograms based on low-level features, can be represented by low-level feature descriptors. These automatically extractable descriptors, such as dominant color and motion trajectory, are suitable for a limited range of applications only (e.g., machine learning-based classification) and are not connected directly to sophisticated human-interpretable concepts, such as concepts depicted in a video, which can be described using high-level descriptors only.

To narrow the semantic gap, feature extraction and analysis can be complemented by machine-interpretable background knowledge formalized with description logics (DL) and implemented in ontology languages, in particular the Web Ontology Language (OWL).

While many knowledge representations employ general-purpose DL, multimedia descriptions, such as the description of moving objects in videos, may utilize spatial, temporal, and fuzzy DL as well. This enables the representation of topological relationships between spatial objects, the representation of the process state and state sequences, quantitative and qualitative spatial reasoning, handling imprecision and uncertainty, and selecting the most probable interpretation of a scene.

The depicted concepts and their relationships are usually described in the Resource Description Framework (RDF), which can express machine-readable statements in the form of subject-predicate-object triples (RDF triples), e.g., scene-depicts-person. The formal definition of the depicted concepts and relationships are derived from controlled vocabularies, ontologies, common sense knowledge bases, and Linked Open Data (LOD) datasets. The regions of interest (ROIs) can be annotated using spatial DL formalisms and media fragment identifiers. The temporal annotation of actions and video events can be performed using temporal

DL and rule-based mechanisms. The fusion of these annotations, including descriptors of different modalities, is suitable even for the machine-interpretable spatiotemporal annotation of complex video scenes.

Based on these structured annotations, various inference tasks can be performed to enable the automated interpretation of images, 3D models, audio contents, and video scenes. For example, video frame interpretation can be performed via abductive reasoning and video event recognition via reasoning over temporal DL axioms. The structured annotation of multimedia contents can be efficiently queried manually and programmatically using the very powerful query language SPARQL, although high-level concept mapping usually requires human supervision and judgment, and automatic annotation options need more research. Research results for high-level concept mapping in constrained videos, such as medical, news, and sport videos, are already promising. The structured description of multimedia contents and the semantic enrichment of multimedia metadata can be applied in video understanding, content-based video indexing and retrieval, automated subtitle generation, clinical decision support, and automated music and movie recommendation engines.

High-level concept mapping relies on formal concept definitions provided by RDFS (RDF Schema) vocabularies and OWL ontologies. However, it is a common and bad practice to create OWL ontologies using a tree structure of concept hierarchies visualized by the graphical user interface of ontology editors, and in particular that of Protégé, without formal grounding in DL.

Without logical underpinning, the computational properties of ontologies remain unclear, and reasoning over RDF statements leveraging formal definitions of ontologies that lack formal grounding might be not decidable. Beyond decidability, a crucial design principle in ontology engineering is to establish favorable trade-offs between expressivity and scalability, and when needed maximize expressivity. However, the higher the DL expressivity, the higher the reasoning complexity. Since the best balance between language expressivity and reasoning complexity depends on the intended application, a variety of DL have been developed, supporting different sets of mathematical constructors.

The level of semantic representation and logical formalization, together with the knowledge base size, presence or absence of instances, and the capabilities of the reasoner, determines what kind of reasoning is feasible. The reasoning tasks utilize different sets of reasoning rules that provide semantics for RDF and RDFS vocabularies, RDFS datatypes, and OWL ontologies, to infer new statements, check knowledge base consistency, determine concept satisfiability, calculate the subsumption hierarchy, and perform instance checking.

The formal knowledge representation of, and reasoning over, depicted objects and events can be used in high-level scene interpretation. Application areas include, but are not limited to, classification, video surveillance, intelligent video analysis, moving object detection and tracking, human intention detection, real-time activity monitoring, next-generation multimedia indexing, and content-based multimedia retrieval.

This book introduces researchers to multimedia semantics by providing an in-depth review of state-of-the-art standards, technologies, ontologies, and software tools. It draws attention to the importance of formal grounding in the knowledge representation of multimedia objects, and the potential of multimedia reasoning in intelligent multimedia applications. It presents both theoretical discussions and multimedia ontology engineering best practices. In this book, the reader familiar with mathematical logic, Internet, and multimedia fundamentals can learn to develop formally grounded multimedia ontologies and map concept definitions to high-level descriptors. The core reasoning tasks, reasoning algorithms, and industry-leading reasoners are also presented, and scene interpretation via reasoning demonstrated. This book aims to illustrate how to use DL-based formalisms to their full potential in the creation, indexing, and reuse of multimedia semantics.

Adelaide, Australia Leslie F. Sikos

Contents

About the Author

Leslie F. Sikos, Ph.D., is a researcher at Flinders University, Australia, specializing in knowledge representation of multimedia resources, multimedia ontology engineering, and automated video scene interpretation via spatiotemporal reasoning and information fusion. He has worked in both academia and the industry, thereby acquiring hands-on skills in Semantic Web technologies, image processing, video authoring, CGI, and 3D modeling. Dr. Sikos is an experienced author with a strong textbook publishing background, who regularly works with the world's leading publishing houses. He has developed two of the most expressive multimedia ontologies to date, which are formally grounded in description logics and complemented by rulesets. Dr. Sikos is a recognized expert in the standardization of next-generation video indexing techniques that leverage semantic annotation at different levels of granularity, including entire videos, video shots, video segments, video frames, and regions of interest of video frames, and mapping depicted concepts to Linked Data. Inspired by the creation and exploitation of rich LOD datasets, Dr. Sikos actively contributes to the development of open standards and open data repositories. For more information, visit https://www.lesliesikos.com.

Springer

http://www.springer.com/978-3-319-54065-8

Description Logics in Multimedia Reasoning
Sikos, L.F.
2017, XIII, 205 p. 25 illus., 20 illus. in color., Hardcover
ISBN: 978-3-319-54065-8