Learning Knowledge Bases for Text and Multimedia

Lexing Xie
Australian National University, NICTA
Canberra, Australia
lexing.xie@anu.edu.au

Haixun Wang
Google Research
Mountain View, CA, USA
haixun@google.com

ABSTRACT

Knowledge acquisition, representation, and reasoning has been one of the long-standing challenges in artificial intelligence and related application areas. Only in the past few years, massive amounts of structured and semi-structured data that directly or indirectly encode human knowledge became widely available, turning the knowledge representation problems into a computational grand challenge with feasible solutions in sight. The research and development on knowledge bases is becoming a lively fusion area among web information extraction, machine learning, databases and information retrieval, with knowledge over images and multimedia emerging as another new frontier of representation and acquisition. This tutorial aims to present a gentle overview of knowledge bases on text and multimedia, including representation, acquisition, and inference. The content of this tutorial are intended for surveying the field, as well as for educating practitioners and aspiring researchers.

Category and Subject Descriptors I.2.6 Artificial Intelligence Learning — Knowledge acquisition

Keywords knowledge graph, multimedia, learning.

1. A MOTIVATING EXAMPLE

On Mar 24, 2014, families were notified by Malaysia Airline that Flight 370 crashed in the southern India ocean, and all lives onboard were lost. The New York Times ran a headline about the tragedy, but unfortunately, the story was juxtaposed with an iPad Air ad, which ends in an underwater scene. The Times pulled the ad after it was notified by a number of readers.

Glitches like this occurr frequently in computational advertising. To prevent it, machines need to understand not just the text but also the image, and then associate the semantics therein to sentiment analysis. It is clear that knowledge is needed in order to process noisy, ambiguous, and scarce input of heterogenous types. The challenges lie in every aspect of knowledge acquisition, representation, and inferencing. Although we are not yet at a stage of completely solving problems such as the one in computational advertisement mentioned above, much recent progress has been made on knowledge intensive text and multimedia processing. This tutorial will survey recent advances on this topic, and discuss fundamental problems, techniques as well as open issues in this vibrant area.

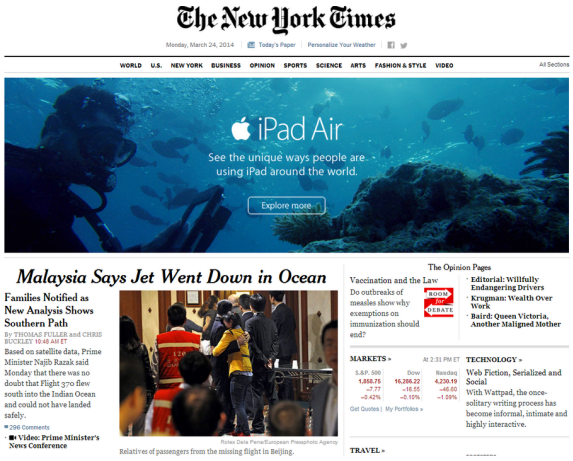

Figure 1: A snapshot of the New York Time homepage, with an iPad Air ad juxtaposed with news about flight MH370.

2. TUTORIAL OVERVIEW

This tutorial is structured in three main parts. The first part introduces the background and foundations of knowledge bases, the second part focuses on knowledge acquisition from text and hyper-text data, the third part discusses acquiring and applying knowledge for image and multimedia collections.

Introduction and foundations

- Beyond search: the age of structure-rich knowledge bases. We will introduce the challenge of learning knowledge bases, and its various applications, in order to motivate and inspire the audience of this problem area. This section will also provide a quick overview for the rest of the tutorial.
- A brief history of knowledge bases. We will present a brief historic note contracting knowledge bases from the mid-1980s, from Cyc [8], Wordnet [11], to other manually curated sources such as ConceptNet [10].
- Introduction to knowledge representation. We will discuss the basic representation of knowledge as multi-relational graphs, and compare-and-contrast example

MM'14, November 3–7, 2014, Orlando, Florida, USA.
ACM 978-1-4503-3063-3/14/11.
http://dx.doi.org/10.1145/2647868.2654851 .

contemporary knowledge bases that possess rich node structure, rich edge information, and their different uses.

- The cognitive science of knowledge structure and representation. This part will present relevant literature in cognitive science about the foundation of knowledge structure. We will discuss Rosch's theory of categorization, as well as work in the macroscopic structure of semantic networks [2], and predicting human memory search [6].

Knowledge acquisition, representation and inference for text

Language is the primary medium for encoding and disseminating knowledge. The representation of real-world knowledge in multimedia is tightly connected to, and often grounded in textual representations. This section covers fundamental and frontier techniques web-scale knowledge bases, and paves the way for discussing research challenges on multimedia knowledge.

- Information extraction from web-scale text. We will introduce systems such as YAGO [14], NELL [3], ReVerb [5], Probase [15] that focus on acquiring knowledge from text. We will also discuss core challenges such as bias from syntactic patterns, and how to evaluate the quality of acquired knowledge.
- Building a data driven semantic network. We will discuss how to score the data using measures such as typicality, basic level of categorization, similarity.
- The conceptualization mechanism. We will discuss how inferencing using the semantic network benefits semantic text processing.
- Applications of knowledge graphs.

Knowledge acquisition and applications for multimedia

- Knowledge and relations in computer vision and image understanding. We will discuss a set of diverse set of approaches to represent and acquire knowledge in computer vision, including Visipedia [12], attribute-based visual recognition, and image understanding with knowledge from WordNet.
- Learning knowledge graphs for multimedia. We will discuss a set of diverse set of approaches for constructing knowledge graphs from images and multimedia, from learning relationships within the visual modality [4], to grounding a textual databases to image instances, to learning knowledge graphs from parallel corpora of media and text.
- Embedding and recommendation models for relationship learning. We take a closer look at several mathematical models for learning knowledge relations from rich media, from relation learning as recommendation [13], to inverse random walk approaches [16].
- Application of multimedia knowledge. We will discuss classical and emerging applications, such as search [1] natural language generation[7], video question-answering [9], and others.

We conclude the tutorial with a view to the open research and application challenges. Tutorial materials and an annotated bibliography will be released online. The audience of this tutorial will take away in-depth knowledge on the construction and application of knowledge graphs, acquire a crucial understanding in the technical challenges in designing and using a knowledge graph, and develop an appreciation to state-of-the-art tools and techniques at the frontiers of this area.

Acknowledgments NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program. This work was supported in part by the Australian Research Council under project DP140102185.

3. REFERENCES

[1] A. B. Benitez and S.-F. Chang. Image classification using multimedia knowledge networks. In *ICIP 2003*, volume 3, pages III–613. IEEE, 2003.

[2] J. Borge-Holthoefer and A. Arenas. Semantic networks: Structure and dynamics. *Entropy*, 12(5):1264–1302, 2010.

[3] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka Jr, and T. M. Mitchell. Toward an architecture for never-ending language learning. In *AAAI*, 2010.

[4] X. Chen, A. Shrivastava, and A. Gupta. Neil: Extracting visual knowledge from web data. ICCV, 2013.

[5] O. Etzioni, M. Banko, S. Soderland, and D. S. Weld. Open information extraction from the web. *Communications of the ACM*, 51(12):68–74, 2008.

[6] T. L. Griffiths, M. Steyvers, and A. Firl. Google and the mind: Predicting fluency with pagerank. *Psychological Science*, 18(12):1069–1076, 2007.

[7] G. Kulkarni, V. Premraj, S. Dhar, S. Li, Y. Choi, A. C. Berg, and T. L. Berg. Baby talk: Understanding and generating simple image descriptions. In *CVPR*, pages 1601–1608. IEEE, 2011.

[8] D. B. Lenat. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, 1995.

[9] G. Li, Z. Ming, H. Li, and T.-S. Chua. Video reference: question answering on youtube. MM '09, pages 773–776, 2009.

[10] H. Liu and P. Singh. ConceptNet – a practical commonsense reasoning tool-kit. *BT technology journal*, 22(4):211–226, 2004.

[11] G. Miller. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.

[12] P. Perona. Vision of a visipedia. *Proceedings of the IEEE*, 98(8):1526–1534, 2010.

[13] S. Riedel, L. Yao, A. Mccallum, and B. M. Marlin. Relation Extraction with Matrix Factorization and Universal Schemas. In *HLT-NAACL '13*, 2013.

[14] F. M. Suchanek, G. Kasneci, and G. Weikum. Yago: A Core of Semantic Knowledge. In *WWW '07*, 2007.

[15] W. Wu, H. Li, H. Wang, and K. Q. Zhu. Probase: a probabilistic taxonomy for text understanding. SIGMOD, 2012.

[16] L. Xie and X. He. Picture tags and world knowledge: Learning tag relations from visual semantic sources. In *ACM Multimedia*, October 2013.