Yu-xin Peng, Wen-wu Zhu, Yao Zhao, Chang-sheng Xu, Qing-ming Huang, Han-qing Lu, Qing-hua Zheng, Tie-jun Huang, and Wen Gao, 2017. Cross-media analysis and reasoning: advances and directions. *Frontiers of Information Technology & Electronic Engineering*, **18**(1):44-57.
http://dx.doi.org/10.1631/FITEE.1601787

Cross-media analysis and reasoning: advances and directions

Key words: Cross-media analysis; Cross-media reasoning; Cross-media applications

Corresponding author: Wen-wu Zhu
E-mail: wwzhu@tsinghua.edu.cn
ORCID: http://orcid.org/0000-0001-7658-3845

Motivation

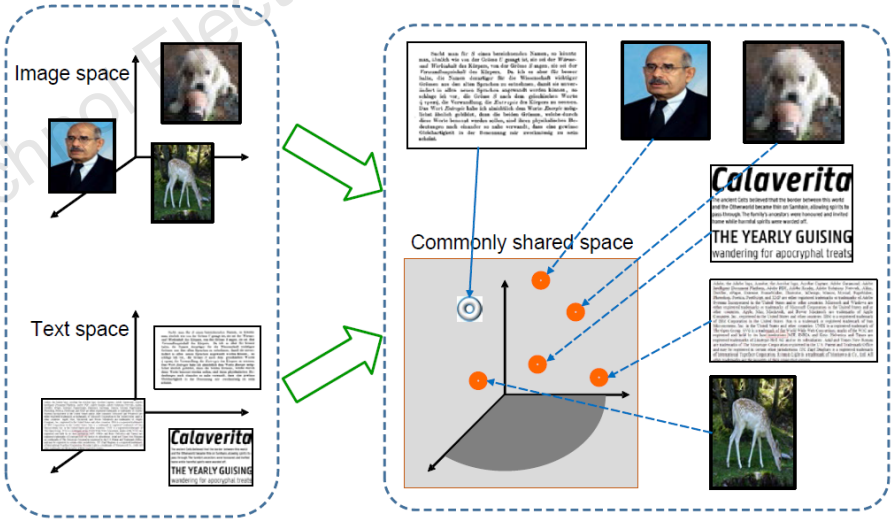
- Cross-media analysis and reasoning is an active research area in computer science, and a promising direction for artificial intelligence
- No existing work has summarized the state-of-the-art methods, or presented advances, challenges, and future directions
- Our goal is not only to present the state-of-the-art advances in the community, but also to provide technical insights by discussing the challenges and research directions

Advances and Directions

- Summarized as ***seven parts***:
  - Theory and model for cross-media uniform representation
  - Cross-media correlation understanding and deep mining
  - Cross-media knowledge graph construction and learning methodologies
  - Cross-media knowledge evolution and reasoning
  - Cross-media description and generation
  - Cross-media intelligent engines
  - Cross-media intelligent applications

Advances and Directions

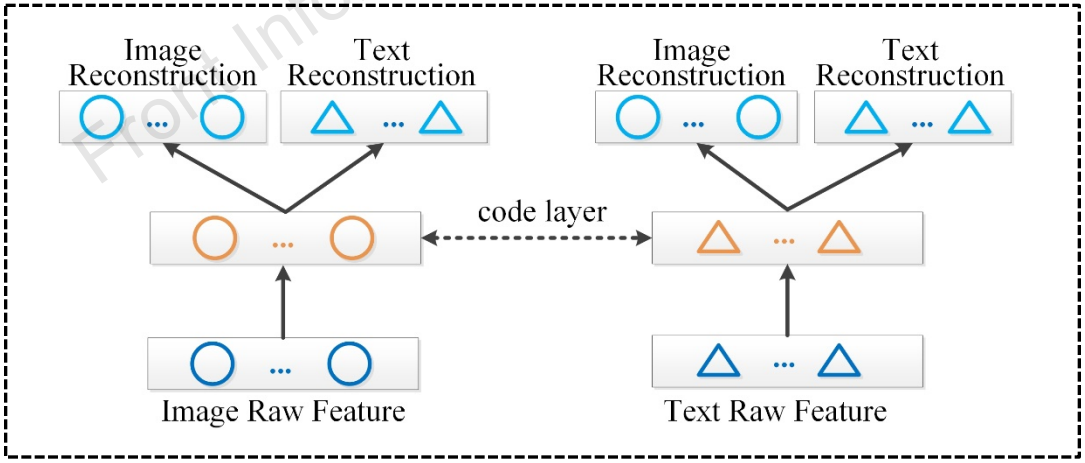
- Theory and model for cross-media uniform representation
  - Aims to build a commonly shared space where cross-media similarities can be computed directly using common distance metrics
  - The key issue is to improve representation accuracy of the existing schemes to meet user expectations



An example of cross-media uniform representation

Advances and Directions

- Cross-media correlation understanding and deep mining
  - Aims to construct metrics on heterogeneous data representation to measure how they are semantically relevant
  - Key issues:
    - Capturing the high-order correlations
    - Adapting to the evolution in heterogeneous entities and relations



A example of correspondence full-modal autoencoder

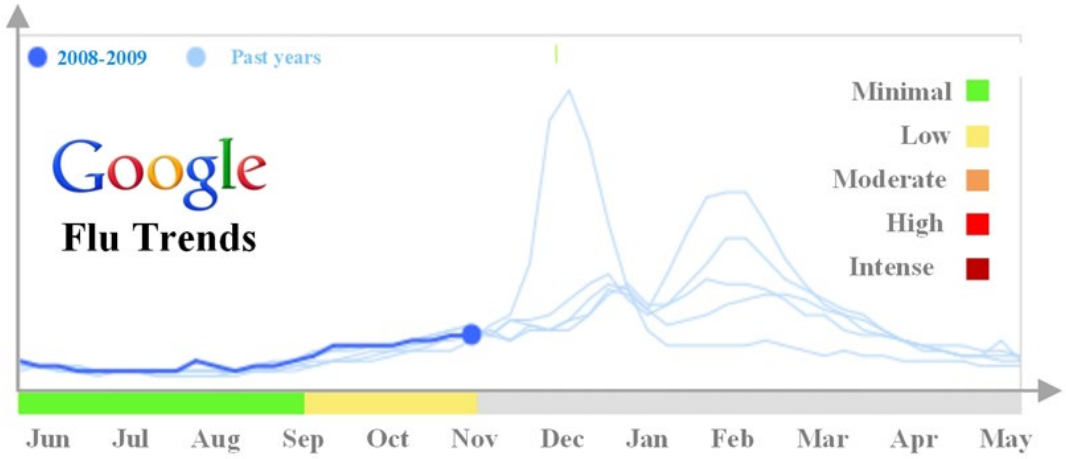
Advances and Directions

- Cross-media knowledge graph construction and learning methodologies
  - Aims to provide essential computable knowledge representation structures, for semantic correlation analysis and cognition-level reasoning in cross-media context
  - Key issues:
    - Cross-media entity extraction and relation construction
    - Information search and retrieval based on cross-media knowledge graphs
    - Mining and reasoning in cross-media knowledge graphs
    - Knowledge-driven cross-media learning models

Advances and Directions

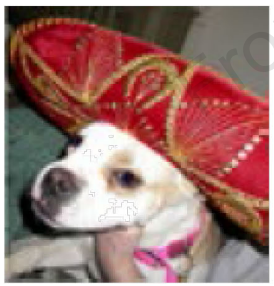
- Cross-media knowledge evolution and reasoning
  - Aims to perform evolution and reasoning processes based on multiple media types rather than on only text
  - Key Issues:
    - Data-driven and knowledge-guided cross-media knowledge learning
    - Cross-media reasoning frameworks based on semantic understanding
    - Never-ending knowledge acquisition, mining, and evolution processes

Google Flu Trends **predicts the spread of influenza by analyzing the search patterns**

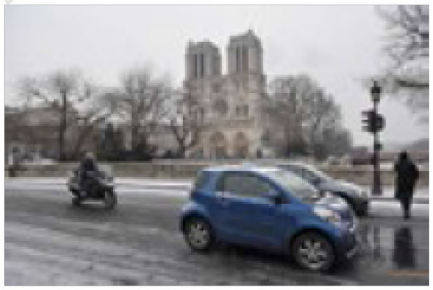


Advances and Directions

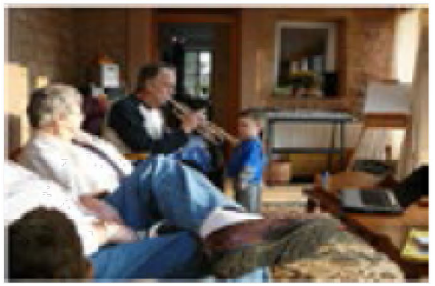
- Cross-media description and generation
  - Realizes cross-translation among cross-media data, and links the multimodal understanding with natural language descriptions
  - Key Issues:
    - Cross-media descriptions of text, image, video, audio, and stereo picture
    - Connections with cognition, emotion, and reasoning



(a) A dog is wearing a red sombrero



(b) Several cars and a motorcycle are on a snow covered street



(c) Some people in chairs and a child watch someone playing a trumpet

Examples of image descriptions

Advances and Directions

- Cross-media intelligent engines
  - Acts as a bridge between technologies and applications, and would be a computing platform for cross-media intelligent applications
  - Key Issues:
    - Adapting to the complex characteristics of cross-media data
    - Coping with the needs of general and specific tasks
    - Building the capacity of autonomic learning and evolution

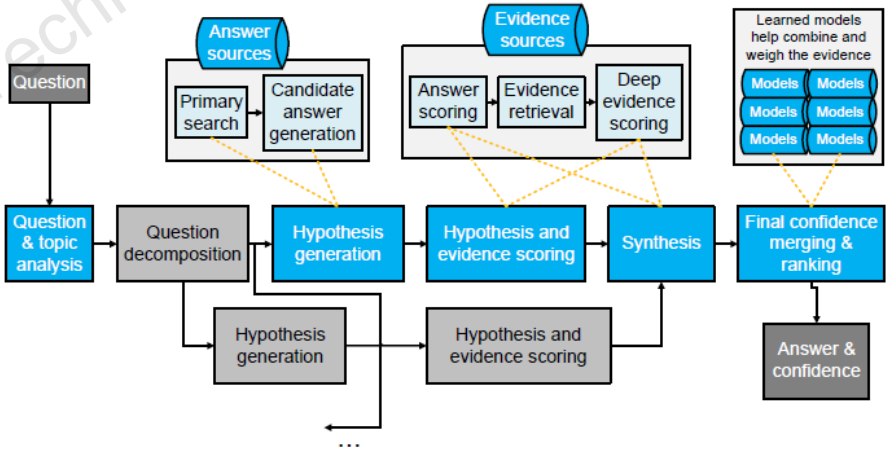


Fig. 7 The high-level architecture of IBM's DeepQA used in Watson

Advances and Directions

- Cross-media intelligent applications
  - Web content monitoring
    - Helps improve the stability of human society
    - Challenges: time lag, insufficient coverage, and high cost
  - Web information trend analysis
    - Aims to alleviate social panic and understand public opinion
    - Challenges: inability to efficiently collect and utilize cross-media data, and ignoring sequential characteristics of public opinion
  - Healthcare data fusion and reasoning
    - Aims to realize personalized and precision medicine
    - Challenges: inability of cross-media fusion and analysis, lack of supervision from domain experts, and poor adaptability toward different medical paradigms

Advances and Directions

- Cross-media intelligent applications

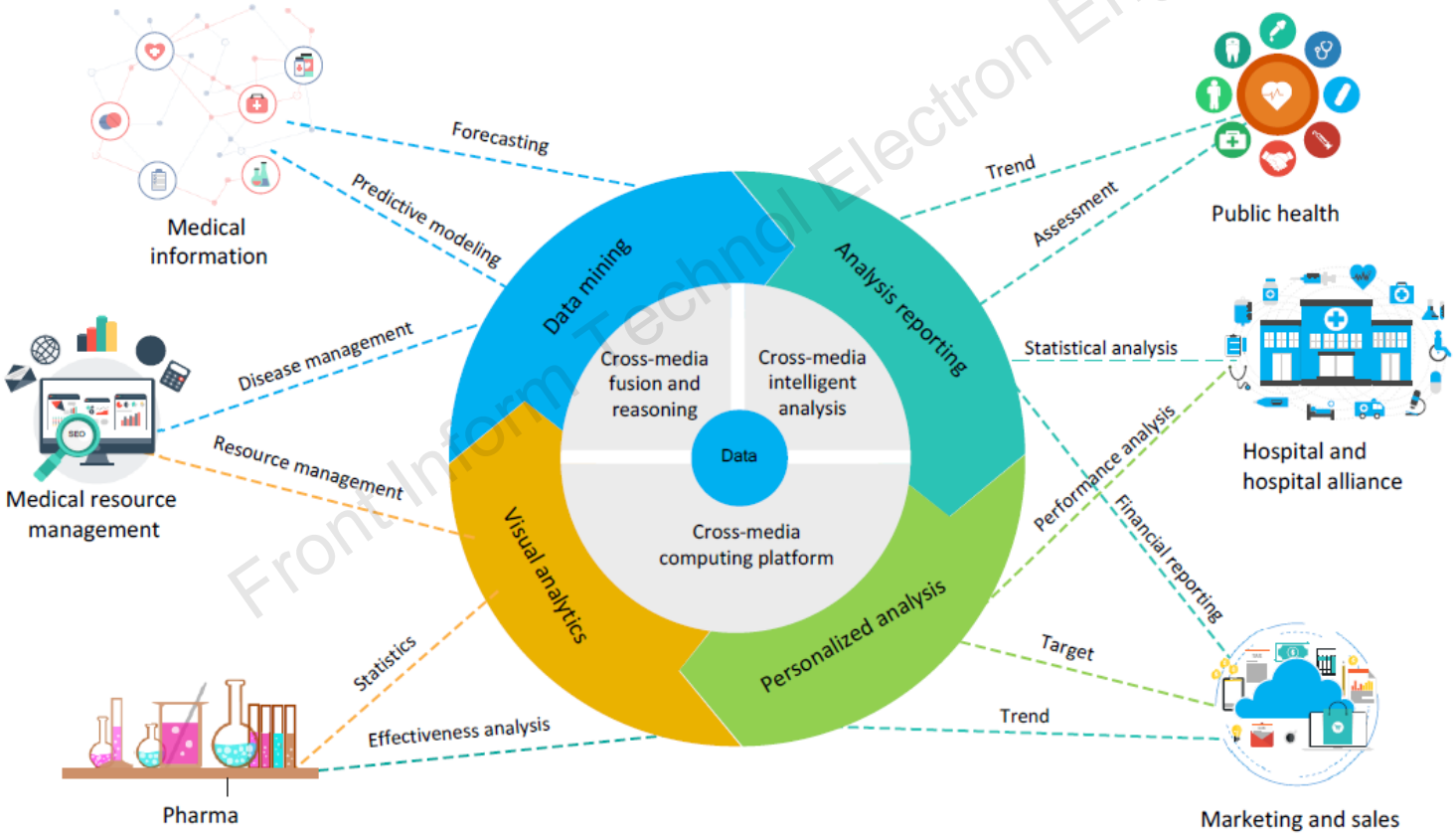


Fig. 9 Existing applications in healthcare analytics

Summary

- This paper has presented the advances achieved by existing studies, as well as the major challenges and open issues
- Cross-media analysis and reasoning has been a key problem of research, and has wide prospects for application
- Although considerable improvement has been made, there remain important challenges in the future research