The quest for large scale adoption of intelligent systems (1/2 day)

Nicolae Duta

Microsoft New England Research and Development Center

One Memorial Drive, Cambridge, MA, 02142, USA

- *A two-sentence description of the tutorial, suitable for inclusion in the conference registration brochure*.

This tutorial is aimed at providing the IJCAI community an overview of some deployed speech, vision and natural language technologies, their historical evolution and the factors which influenced their adoption. We review several fundamental problems and trade-offs encountered when building practical AI systems: deterministic vs stochastic systems, supervised vs unsupervised vs semi-supervised learning systems, client only vs client-server architectures, data annotation challenges, deployment, evaluation and monetization.

- *A two-paragraph description of the tutorial, suitable for a web page overview*.

Scientists have long dreamed of creating machines people could interact with in a humanly way. Although one no longer believes Turing's prophecy that machines will be able to converse like humans in the near future, real progress has been made in vision, voice and text-based technologies and they have become an essential part of many human-machine interaction systems (and even human-to-human: automated translation and speech-to-speech systems). There are now vision and fingerprint personal identification, voice-based personal assistants, search and transactional systems for most smart phone platforms. The technology is pushed even further by the search engines which have evolved from simple keyword search to semantic search (they can now provide direct answers to a wide range of questions) as well as enterprise applications like automated document understanding.

This tutorial is aimed at providing the IJCAI community an overview of some deployed speech, vision and natural language technologies, their historical evolution and the factors which influenced their adoption. We review several fundamental problems and trade-offs encountered when building practical AI systems: deterministic vs stochastic systems, supervised vs unsupervised vs semi-supervised learning systems, client only vs client-server architectures, data annotation challenges, deployment, evaluation and monetization.

- *A detailed outline of the tutorial* (½ day)

    - Introduction and history/examples of deployed large-scale intelligent systems

        - Voice-based: Automated Directory Assistance, Call routing, Directed Dialog, Personal Assistants

        - Text-based: Question answering, Information retrieval, Semantic search

        - Vision-based: Object detection, recognition, matching

        - Text & Vision based: Automated document understanding

        - DARPA programs and evaluations

    - Brief intro to voice and vision-based technologies

        - Acoustic modelling & Speech recognizer architecture

        - Language prediction/modelling & understanding

        - Information extraction, question/query answering & semantic search

        - Computer vision

    - AI system design

        - Client only vs. client-server architectures; system components

        - Deterministic vs Statistical/ML vs hybrid systems

        - Supervised vs Unsupervised vs Semi-supervised learning

        - Data annotation challenges

        - Deployment issues and evaluation

    - Factors contributing to AI system adoption

        - System Accuracy

            - What is acceptable accuracy in production?

            - What if the accuracy is not sufficiently high?

        - System Latency

            - Sources of latency

        - System reliability

- - Internal and external contributing factors
  - System maintenance
    - Deployment on premise vs in the cloud
    - Handling technical debt
  - Costs
    - Development costs, usage costs, maintenance costs
  - System monetization and benefits for the user/customer
    - o Future developments
- *A characterization of the potential target audience for the tutorial, including prerequisite knowledge.*

Potential target audience of the tutorial: AI/Machine learning researchers who would like an introduction to enterprise voice & vision technologies as well as practical considerations on building, deploying and monetizing such systems

Prerequisite knowledge: Elementary probability/statistics. Some pattern recognition & machine learning knowledge may also be useful.

The following objectives are best served by the proposed tutorial:

Introduce expert non-specialists to an AI subarea (voice & vision technologies).
Motivate and explain a topic of emerging importance for AI.

- *A description of why the tutorial topic would be of interest to a substantial part of the IJCAI audience.*

The last decade brought a proliferation of smart phones and wearable devices along with applications that can be driven by voice and vision. They not only recognize the person and the words being said but also their meaning and are many times able to give answers back or ask for additional information. Such systems are based on technologies from many AI domains: machine learning, knowledge representation, reasoning and retrieval, human-machine interfaces, etc. The proposed tutorial describes how theories and methods from these AI domains can be used to build intelligent systems and what factors influence their adoption.

- *Brief resume of the presenter*:

Nicolae Duta

Microsoft New England Research and Development Center

One Memorial Drive, Cambridge, MA, 02142, USA

Tel: (857) 4536243

niduta@microsoft.com

Nicolae Duta received the B.S. degree in applied mathematics from the University of Bucharest (Romania) in 1991, the D.E.A. degree in statistics from the University of Paris-Sud (France) in 1992, the M.S. degree in computer science from the University of Iowa in 1996 and the Ph.D. degree in computer science and engineering from Michigan State University in 2000. He is currently a senior scientist in the Applied AI Group at Microsoft in Cambridge, MA working on query understanding for Bing and Cortana personal assistant. From 2006 to 2013 he was a member of the Natural Language Understanding and Language Modeling groups at Nuance Communications, Burlington, MA where he developed Dragon Go - the first generation of voice-based personal assistants for smart phones. From 2000 to 2005 he was a scientist in the Speech and Language Processing department at BBN Technologies, Cambridge, MA. He also held temporary research positions at INRIA-Rocquecourt (France) in 1993 and Siemens Corporate Research (Princeton, NJ) from 1997 to 1999. He is a member of IEEE and his current research interests include computer vision, pattern recognition, language understanding, automatic translation, machine and biological learning.

Background in the tutorial area: 25 years of research in Computer vision and Speech and Language Technologies at Siemens Research, BBN Technologies, Nuance Communications and Microsoft. Authored 4 US/international patents and 3 pending patent applications in vision and language technologies.

Citation to an available example of work in the area-- ideally, a published tutorial-level article on the subject

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Francois Crespo, and Dan Dennison. 2015. Hidden technical debt in Machine learning systems. In Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'15), C. Cortes, D. D. Lee, M. Sugiyama, and R. Garnett (Eds.), Vol. 2. MIT Press, Cambridge, MA, USA, 2503-2511.

Nicolae Duta: Natural Language Understanding and Prediction: from Formal Grammars to Large Scale Machine Learning, Fundamenta Informaticae, Volume 131(3), 2014, pp. 425-440