Emerging Topics in Learning from Noisy and Missing Data

Emerging Topics in Learning from Noisy and Missing Data

Xavier Alameda-Pineda
University of Trento
Trento, Italy
xavier.alamedapineda@unitn.it

Timothy M. Hospedales
Queen Mary University
London, UK
t.hospedales@qmul.ac.uk

Elisa Ricci
FBK / Univ. of Perugia
Trento, Italy / Perugia, Italy
eliricci@fbk.eu

Nicu Sebe
University of Trento
Trento, Italy
sebe@disi.unitn.it

Xiaogang Wang
The Chinese University of Hong Kong
Hong Kong, China
xgwang@ee.cuhk.edu.hk

ABSTRACT

While vital for handling most multimedia and computer vision problems, collecting large scale fully annotated datasets is a resource-consuming, often unaffordable task. Indeed, on the one hand datasets need to be large and variate enough so that learning strategies can successfully exploit the variability inherently present in real data, but on the other hand they should be small enough so that they can be fully annotated at a reasonable cost. With the overwhelming success of (deep) learning methods, the traditional problem of balancing between dataset dimensions and resources needed for annotations became a full-fledged dilemma. In this context, methodological approaches able to deal with partially described data sets represent a one-of-a-kind opportunity to find the right balance between data variability and resource-consumption in annotation. These include methods able to deal with noisy, weak or partial annotations. In this tutorial we will present several recent methodologies addressing different visual tasks under the assumption of noisy, weakly annotated data sets.

Keywords

Noisy and missing data; Low rank models; Zero-shot learning; Deep learning

1. COURSE DESCRIPTION

In the last decades the bond between machine learning, computer vision and multimedia has strengthened and a large variety of powerful learning frameworks have emerged for the automatic analysis of visual data. Earlier works addressed visual tasks considering the ideal condition in which features were complete and labels were available for all training samples. The research community soon realized that acquiring fully described data sets for certain tasks was resource-consuming, and often not affordable. Exemplar tasks are classification of web images, social signal processing in unrestricted environments, face analysis in the wild or the recognition of subjective properties of images and videos. In the last few years, the success of data-hungry deep architectures has made this problem even more evident. Indeed, data sets need to be large and variate enough so that neural network-based learning strategies can successfully exploit the variability inherently present in real data, and small enough so that they can be fully annotated at a reasonable cost. In this context, methodological approaches able to deal with noisy, weak or partial annotations are of particular interest.

In this tutorial we will present several recent methodologies addressing different visual tasks under the assumption of noisy, weakly annotated data sets. Special emphasis will be given to methods based on deep architectures for unsupervised domain adaptation [6, 5], low-rank modeling for learning in transductive settings and zero-shot learning [3]. We will show how these approaches exhibit excellent performance in crucial tasks such as pedestrian detection or fine-grained visual recognition. Furthermore, we will discuss emerging application domains which are of great interest to the multimedia community and where handling noisy or missing information is essential. For instance, we will present recent works on multimodal complex scene analysis using wearable sensors [2], on the estimation of physiological signals from face videos in realistic conditions [4], and on the recognition of emotions elicited from abstract paintings [1].

1.1 Outline

The total duration of the tutorial is 3 hours. The topics to be covered (including estimated duration of each topic) are:

- Introduction: challenges in learning from noisy and missing data (30 min)
- Robust low-Rank modeling and transductive learning in emerging application domains (60 min)
- Zero-shot learning (45 min)
- Unsupervised domain adaptation in deep neural architectures (45 min)

MM '16 October 15-19, 2016, Amsterdam, Netherlands

ACM ISBN 978-1-4503-3603-1/16/10...$15.00

DOI: http://dx.doi.org/10.1145/2964284.2986910

2. TARGET AUDIENCE

The intended audience (intermediate/advanced level) is (i) researchers interested in vision and multimedia applications necessitating robust learning approaches and (ii) researchers interested in sparse modelling and deep learning in scarcely labeled setting. Slides of the presentations and code associated to the main algorithms presented in the course will be distributed to the audience.

3. RELATED TUTORIALS

This tutorial will provide an overview of very recent methodologies for addressing an increasingly pressing need in multimedia and computer vision: the possibility to learn robust models from noisy and partially annotated datasets.

This one-of-a-kind course (recent tutorials do not specifically focus on this topic) will be of great interest for the research community, both from the methodological and from the application perspective. Specifically, this tutorial will introduce several algorithms for handling noisy, partially annotated data, focusing on recent learning paradigms, such as low rank modelling, deep learning and domain adaptation. The methodologies that will be presented in this short course (e.g. zero-shot learning, deep domain adaptation and matrix completion) have not been considered in previous tutorials. Furthermore, we will discuss emerging application scenarios which benefit from the proposed techniques and are of great interest for the multimedia community, given recent related tutorials in the areas of social signal processing and human-centric visual content analysis:

- Emotional and Social Signals for Multimedia Research (ACM MM 2015),
- Human-centric images and videos analysis (ACM MM 2015),
- Social Multimedia Computing: A User-centric Perspective (ICME 2015).

4. BIOGRAPHIES

Xavier Alameda-Pineda received his PhD from INRIA and University of Grenoble in 2013. He was a post-doctoral researcher at CNRS/GIPSA-Lab. Currently he holds a research fellowship at the University of Trento, in the Deep Relational Group, working on signal processing and machine learning for scene and behavior understanding using multimodal data. He is the IEEE WASPAA 2015 best student paper award winner and the ACM Multimedia 2015 best paper prize winner.

Timothy Hospedales is a Senior Lecturer (associate professor) at Queen Mary University of London. He leads the Applied Machine Learning lab at QMUL, studying applications of weakly supervised, transfer and cross-modal learning within computer vision and multimedia. He was area chair of WACV 2016 and BMVC 2015 best paper prize winner.

Elisa Ricci is a researcher at FBK and an assistant professor at the University of Perugia. She received her PhD from the University of Perugia in 2008. She has been a post-doctoral researcher at Idiap and FBK, Trento and a visiting researcher at the University of Bristol. Her research interests are directed along developing machine learning algorithms for video scene analysis, human behavior understanding and multimedia content analysis. She is area chair of ACM MM 2016 and of ECCV 2016.

Nicu Sebe is a full professor in the University of Trento, Italy, where he is leading the research in the areas of multimedia information retrieval and human behavior understanding. He was a General Co-Chair of FG 2008 and ACM MM 2013, and a program chair of CIVR 2007 and 2010, and ACM MM 2007 and 2011. He is a program chair of ECCV 2016 and ICCV 2017. He is a senior member of IEEE and ACM and a fellow of IAPR.

Xiaogang Wang received the PhD degree in Computer Science from Massachusetts Institute of Technology in 2009. He is an associate professor in the Department of Electronic Engineering at the Chinese University of Hong Kong. He received the Outstanding Young Researcher in Automatic Human Behavior Analysis Award in 2011, HK RGC Early Career Award in 2012, and Young Researcher Award of the Chinese University of Hong Kong. He was the area chair of ICCV 2011, ICCV 2015, ECCV 2014, 2016 and ACCV 2014. His research interests include computer vision, deep learning, crowd video surveillance, object detection, and face recognition.

5. REFERENCES

[1] X. Alameda-Pineda, E. Ricci, Y. Yan, and N. Sebe. Recognizing emotions from abstract paintings using non-linear matrix completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5240–5248, Las Vegas, USA, 2016.

[2] X. Alameda-Pineda, Y. Yan, E. Ricci, O. Lanz, and N. Sebe. Analyzing free-standing conversational groups: a multimodal approach. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 5–14, Brisbane, Australia, 2015.

[3] Y. Fu, T. M. Hospedales, T. Xiang, and S. Gong. Transductive multi-view zero-shot learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(11):2332–2345, 2015.

[4] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe. Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2396–2404, Las Vegas, USA, 2016.

[5] T. Xiao, T. Xia, Y. Yang, C. Huang, and X. Wang. Learning from massive noisy labeled data for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2691–2699, Boston, USA, 2015.

[6] X. Zeng, W. Ouyang, M. Wang, and X. Wang. Deep learning of scene-specific classifier for pedestrian detection. In *European Conference on Computer Vision*, pages 472–487, Zurich, Switzerland, 2014.