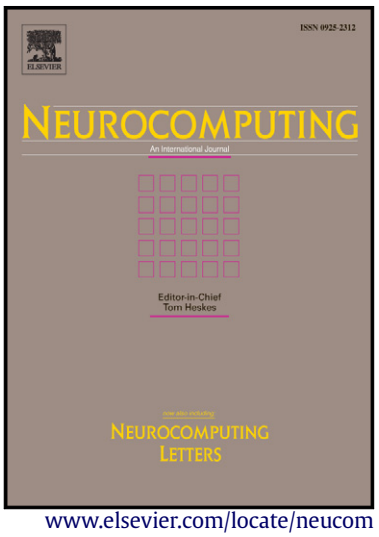
Special issue on advances in learning with label noise

Frénay, Benoît; Kabán, Ata

DOI:
10.1016/j.neucom.2015.01.067



www.elsevier.com/locate/neucom

Special Issue on Advances in Learning with Label Noise

Benoi\widehatt Frénay, Ata Kabán

PII: S0925-2312(15)00140-X
DOI: http://dx.doi.org/10.1016/j.neucom.2015.01.067
Reference: NEUCOM15120

To appear in: *Neurocomputing*

Cite this article as: Benoi\widehatt Frénay, Ata Kabán, Special Issue on Advances in Learning with Label Noise, *Neurocomputing,* http://dx.doi.org/10.1016/j.neucom.2015.01.067

Special Issue on Advances in Learning with Label Noise

Label noise is an important issue in classification. It is both expensive and difficult to obtain completely reliable labels, yet traditional classifiers expect a perfectly labelled training set. In real-world data sets, however, the labels available often contain mistakes. Mislabelling may occur for several reasons, including lack of information, speedy labelling by non-experts, the subjective nature of class memberships, expert errors, and communication problems. Furthermore, label noise may take several different forms – for instance, labelling errors may occur at random, or may depend on particular values of the data features, or they may be adversarial. Errors may affect all data classes equally or asymmetrically. A large body of literature exists on the effects of label noise, which shows that mislabelling may detrimentally affect the classification performance, the complexity of the learned models, and it may impair preprocessing tasks such as feature selection.

Many methods have been proposed to deal with label noise. Filter approaches aim at identifying and removing any mislabelled instances. Label noise sensitive algorithms aim at dealing with label noise during learning, by modelling the process of label corruption as part of modelling the data. Some methods have been modified to take label noise into account in an embedded fashion. The current literature on learning with label noise is a lively mixture of theoretical and experimental studies which clearly demonstrate both the complexity and the importance of the problem. Dealing with mislabelled instances needs to be flexible enough to accommodate label uncertainty, yet constrained enough to guide the learning process regarding when to trust the label and when to trust the classifier.

This special issue contains 13 articles presenting new advances and findings to deal with label noise. In particular, several works propose new methods to deal with label noise and assess them through extensive experiments. Theoretical results about learning in the presence of label noise are also introduced. Label noise is also considered in non-standard settings, like e.g. high-dimensional data and crowd labelling.

New methods to deal with label noise are proposed in 7 papers. ”Comparison of two Topological Approaches to deal with Noisy Labeling” by Rico et al. uses neighbourhood graph to detect potentially mislabelled instances. Two methods are discussed, based on cut edges weighted statistics (CEW) and a relaxation technique (RT). ”Small margin ensembles can be robust to class-label noise” by Sabzevari et al. uses subsampling to generate ensembles that are robust to label noise. ”Correction of Noisy Labels via Mutual Consistency Check” by Bhadraa and Hein proposes a new preprocessing technique for correcting labels by maximising a criterion that predicts the labels of training data based on their neighbours using Parzen window estimators, under the assumption that a small trusted correctly labelled set is available. ”Support Vector Machines Under Adversarial Label Contamination” by Xiaoa et al. evaluates the security of SVMs to mallicious adversarial label noise attacks and proposes an LP relaxation approach to solve the otherwise NP-hard problem. ”Particle Competition and Cooperation for Semi-Supervised Learning with Label Noise” by Brevea et al. presents a robust graph-based method for semi-supervising learning with label noise. ”A Robust Ensemble Approach to Learn From Positive and Unlabeled Data Using SVM Base Models” by Claesen et al. proposes an ensemble technique to learn binary classifiers from a labelled positive class and a pool of unlabelled data by casting it as a classificaton problem with label noise in the negative set. ”Regularized Maximum Correntropy Machine” by Wang et al. presents a non-linear multi-class classifier using an unconventional loss function derived from information theoretic learning, which is robust against label noise.

Theoretical results about learning with label noise are introduced in 4 papers. ”Making Risk Minimization Tolerant to Label Noise” by Ghosh et al. investigates theoretically the robustness of risk minimisation to label noise. It proves a sufficient condition of loss robustness to uniform label noise and analyses the behaviour of several standard losses. ”Effect of label noise in the complexity of classification problems” by Garcia et al. investigates how noise affects the complexity of classification problems. It suggests the use of some data complexity measures (which are more sensitive to label noise) to obtain new filters for noise identification. ”Relating ensemble diversity and performance: A study in class noise detection” by Sluban and Lavrač explores the relationship between ensemble diversity and noise detection for label noise detection with ensembles. ”Accuracy Analysis of Semi-supervised Classification when the Class Balance Changes” by Yamazaki provides an in-depth analysis and new theoretical results on maximum likelihood and Bayesian learning in the semi-supervised setting, and quantifies the loss of accuracy caused by label noise.

Label noise is studied in non-standard settings in 2 papers. ”Modeling Annotator Behaviors for Crowd Labeling” by Kara et al. investigates different annotator behaviours and incorporates them into Bayesian models, since it is more and more common to use crowds to annotate unlabelled data. ”Hubness-aware kNN Classification of High-dimensional Data in Presence of Label Noise” by Tomašev and Buza studies the potiential impact of hubs in presence of label noise.

Benoît Frénay
Université de Namur - Faculty of Computer Science
Rue Grandgagnage 21, 5000 Namur - Belgium
E-mail: benoit.frenay@unamur.be

Ata Kabán
University of Birmingham - School of Computer Science
Edgbaston Birmingham B15 2TT - United Kingdom
E-mail: ata.x.kaban@gmail.com

Acknowledgements

The guest editors would like to thank the Editor-in-Chief Tom Heskes for giving the opportunity to publish this special issue, the Neurocomputing staff for their help throughout the editorial process, the reviewers for their essential reviewing work and the authors for their support by submitting their work.