Selection of Negatives in Hopfield Networks

Marco Frasca
Computer Science Department
University of Milan
Via Comelico 39, 20135 Milan, Italy

In several real-world contexts ranging from text categorization [1] to protein function prediction [2], the notion of “negative” example is not well defined, and the selection of appropriate sets of negative patterns may sensibly improve the predictive capabilities of supervised and semi-supervised methodologies for inductive inference. This area of machine learning has been named Positive-Unlabeled (PU) learning, and there has been a surge of interest lately in this particular subset of semi-supervised learning problems [2, 3]. Indeed, in different contexts there is no "gold standard" for negative items and only positives are the results of accurate studies (negatives usually are simply "non positives").
Several approaches have been proposed for selecting negatives, such as randomly sampling items and assuming the probability of getting a false negative is low [5], sampling according to positive-negative similarity measures [8], selecting the items positive for the sibling and/or ancestral categories of the category of interest as negative examples [6].
Nevertheless, hierarchical methods cannot be applied in context where the categories are not structured as a hierarchy (for instance the churn prediction for service providers [7] and the action video detection [8]). Furthermore, even when parent-child relationships are available for the classes being predicted, strategies based on sibling/ancestral categories may often break down, as some items are annotated to more than one sibling category, and many items have few siblings to use [4].

In this work we propose a novel methodology for graph-based semi-supervised learning which is composed of two main steps: *Step 1*) a novel strategy for PU learning specific for Hopfield networks, which can be applied both to structured classes and to hierarchy-less contexts; *Step 2*) a semi-supervised classifier based on a family of parametric Hopfield networks, which embeds the negative selection performed at Step 1) in the dynamics of network.

At Step 1, the approach for detecting negative instances can be summarized as follows:

1.1) transform the matrix describing symmetric connections between instances into a feature matrix, in which each instance is associated with a 3-feature vector, obtained through the application of 1, 2 and 3 steps random walks from positives. It has been shown that a so obtained feature matrix suffices to propagate information coded in the graph labels [9];
1.2) the positive points are clustered using a dynamic version of the fuzzy C means algorithm [10] exploiting a suitable index [11] in order to decide the

optimal number of clusters to summarize data;

1.3) each negative point is assigned a score consisting in the maximum membership it has to the detected clusters of positives. This ranking of negatives allows to perform an initial discrimination of negatives, by considering negatively labelled those points which are not definitely attributable to any of the clusters (that is with a low value for the computed score);

1.4) the remaining points are then further discriminated according their stability in the Hopfield network constructed from the input graph. In particular, we extend the set of negative with those points that locally minimize the energy function.

At Step 2) the Hopfield network is simulated through an asynchronous dynamics, in which the items not selected during Step 1 are allowed to change their state along with the items whose label has to be predicted (test set). The final equilibrium state is used to infer the prediction for instances in the test set, in particular the instances fired at equilibrium are considered candidates for the positive class.

We experimentally validated the proposed methodology in the protein function prediction problem, which consists in inferring the biomolecular functions of recently discovered proteins starting from their interactions with already characterized proteins, and in which, apart rare cases, just the positive annotations are available for the Gene Ontology classes [12]. The comparison with the state-of-the-art supervised and semi-supervised label prediction methods and with the “vanilla” Hopfield network shows the effectiveness of our approach. Moreover, the reduced computational complexity, due to the application of clustering techniques only to positive instances (which usually are a large minority of the data set), allows the application of the proposed methodology to contexts characterized by large-size data.

References

[1] . T. Fagni , F. Sebastiani (2007). On the selection of negative examples for hierarchical text categorization. *In Proceedings of The 3rd Language Technology Conference*, pages 24-28.

[2] E Ferretti et al (2014). On the Use of Reliable-Negatives Selection Strategies in the PU Learning Approach for Quality Flaws Prediction in Wikipedia. *In Proceedings of the 25th International Workshop on Database and Expert Systems Applications (DEXA’14): 11th International Workshop on Text-based Information Retrieval (TIR’14).*

[3] X. Wang et al. (2010). Semi-supervised Learning from only Positive and Unlabeled Data using Entropy. *Lecture Notes in Computer Science,* **6184**, pages 668-679 .

[4] N. Youngs, D. P. Brown, K. Drew, D. Shasha R. Bonneau (2013). Parametric Bayesian priors and better choice of negative examples improve protein function prediction. *Bioinformatics***, 29,** 1190-1198.

[5]. S. M. Gomez et al. (2003) Learning to predict protein- protein interactions. *Bioinformatics*, **19**, 1875–1881.

[6]. S. Mostafavi Q. Morris (2009). Using the Gene Ontology hierarchy when predicting gene function. AUAI Press. pp. 419–427.

[7] J. Hadden, A. Tiwari, R. Roy, and D. Ruta (2006). Churn prediction using complaints data. *In Proceedings Of World Academy Of Science, Engineering and Technology*.

[8] G.J. Burghouts, K. Schutte, H. Bouma, R.J.M. den Hollander (2014). Selection of negative samples and two-stage combination of multiple features for action detection in thousands of videos. *Machine Vision and Applications*, ***25***: pages 85-9a8.

[9] S. Mostafavi, A. Goldenberg, Q. Morris (2012). Labeling nodes using three degrees of propagation. *PLoS One*. **7**(12):e51947

[10] J.C. Bezdek, R. Ehrlich, W. Full (1984). FCM: The fuzzy c-means clustering algorithm, Computers & Geosciences 19 (2-3), 191-203.

[11] R.J.G.B. Campello, E.R. Hruschka (2006). A fuzzy extension of the silhouette width criterion for cluster analysis, Fuzzy Sets and Systems, 157 (21), pp 2858-2875.

[12] M. Ashburner et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics **25**(1), 25–29.