Using Bayesian networks to predict changes

Sarah Nadi SNADI@UWATERLOO.CA
University of Waterloo, Waterloo, ON, Canada

Table 1. Classification accuracies for naive Bayes and flexible Bayes on various data sets.

DATA SET	NAIVE	FLEXIBLE	BETTER?
BREAST	95.9± 0.2	96.7± 0.2	√
CLEVELAND	83.3± 0.6	80.0± 0.6	×
GLASS2	61.9± 1.4	83.8± 0.7	√
CREDIT	74.8± 0.5	78.3± 0.6	
HORSE	73.3± 0.9	69.7± 1.0	×
META	67.1± 0.6	76.5± 0.5	√
PIMA	75.1± 0.6	73.9± 0.5	
VEHICLE	44.9± 0.6	61.5± 0.4	√

Abstract

STILLLL

1. Introduction

Changes

2. Related Work

Mirarab et al. (Mirarab et al., 2007) investigate the same problem as our work. However, their work is on the level of source code changes. They build three different Bayesian Networks, one that is based on package and class dependency information (static relationships), one which is dependent on historical co-changes, and one which uses both. For the first graph, the initial structure is essentially "given" according to the static dependencies, and then the CPTs are learnt using the importance sampling algorithm proposed by Changhe and Marek (Yuan & Druzdzel, 2003). The way static dependencies are defined in their case is specific to Java. The third one is essentially the first graph, but updated using the historic change information according to the Expectation Maximization (EM) algorithm (Dempster et al., 1977). The second was solely based on historic information where the network is build using a greedy structure learning algorithm (Friedman & Goldzsmidt, 1996). They did some preprocessing to their data such as filtering out large changes (with more than 30 elements changed at once) since this was probably an insignificant change.

Zhou

Algorithm 1 Bubble Sort

Input: data x_i, size m
repeat
  Initialize $noChange = true$.
  for $i = 1$ **to** $m - 1$ **do**
    if $x_i > x_{i+1}$ **then**
      Swap x_i and x_{i+1}
      $noChange = false$
    end if
  end for
until $noChange$ is $true$

3. Building the model

4. Experimental Work

5. Conclusion

Appearing in *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel, 2010.

References

Dempster, A.P., Laird, N.M., Rubin, D.B., et al. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977. ISSN 0035-9246.

Friedman, N. and Goldzsmidt, M. Learning Bayesian Networks with Local Structure. 1996.

Mirarab, S., Hassouna, A., and Tahvildari, L. Using bayesian belief networks to predict change propagation in software systems. pp. 177 –188, jun. 2007.

Yuan, C. and Druzdzel, M.J. An importance sampling algorithm based on evidence pre-propagation.

In *Proceedings of the 19th Annual Conference on Uncertainty on Artificial Intelligence*, pp. 624–631. Citeseer, 2003.