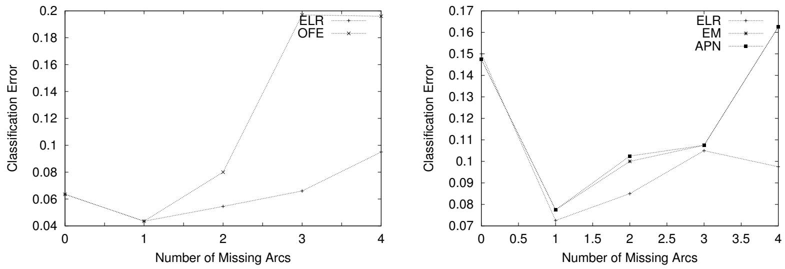

Figure 1: "Correctness of Structure": Comparing ELR to OFE, on increasingly incorrect structures for (a) Complete Data; (b) Incomplete Data

5.1.4 "Correctness of Structure" Study

The NaïveBayes-assumption, that the attributes are independent given the classification variable, is typically incorrect. This is known to handicap the NaïveBayes classifier in the standard OFE situation; see above and [DP96].

We saw above that ELR is more robust than OFE, in that it is not as handicapped by an incorrect structure. We designed the following simple experiment to empirically investigate this claim.

We used synthesized data, to allow us to vary the "incorrectness" of the structure. Here, we consider an underlying distribution P_0 over the $k+1$ binary variables $\{C, E_1, E_2, \ldots, E_k\}$ where (initially) we made NaïveBayes-assumptions and set[1]

$$P(+c) = 0.9 \qquad P(+e_i \mid +c) = 0.2 \qquad P(+e_i \mid -c) = 0.8 \tag{1}$$

and our queries were all complete; *i.e.*, each instance of the form $\mathbf{E} = \langle \pm e_1,\ \pm e_2,\ \ldots,\ \pm e_k \rangle$.

We then used OFE (resp., ELR) to learn the parameters for the NaïveBayes structure from a data sample, then used the resulting BN to classify additional data. As the structure was correct for this P_0 distribution, both OFE and ELR did quite well, efficiently converging to the optimal classification error.

We then tried to learn the CPtables for this NaïveBayes structure, but for distributions that were *not* consistent with this structure. In particular, we formed the m-th distribution P_m by asserting that $E_1 \equiv E_2 \equiv \ldots \equiv E_m$ (*i.e.*, $P(+e_i \mid +e_1) = 1.0$, $P(+e_i \mid -e_1) = 0.0$ for each $i = 2..m$) in addition to Equation 1. Hence, P_0 corresponds to the $m = 0$ case. For $m > 0$, however, the m-th distribution cannot be modeled as a NaïveBayes structure, but could be modeled using that structure augmented with $m-1$ links, connecting E_{i-1} to E_i for each $i = 2..m$.

Figure 1(a) shows the results, for $k = 5$, based on 400 instances. As predicted, ELR can produce reasonably accurate CPtables here, even for increasingly wrong structures. However, OFE does progressively worse.

"Correctness of Structure", Incomplete Data: We next degraded this training data by randomly removing the value of each attribute, within each instance, with probability 0.5. Figure 1(b) compares ELR with the standard systems APN and EM; again we see that ELR is more accurate, in each case.

References

[DP96] P. Domingo and M. Pazzani. Beyond independence: conditions for the optimality of the simple Bayesian classier. In *Proc. 13th International Conference on Machine Learning*, 1996.

[1] For binary variables, we let "$+c$" represent $c =$ True, and "$-c$" represent $c =$ False.