



Ensemble Pruning via Individual Contribution Ordering

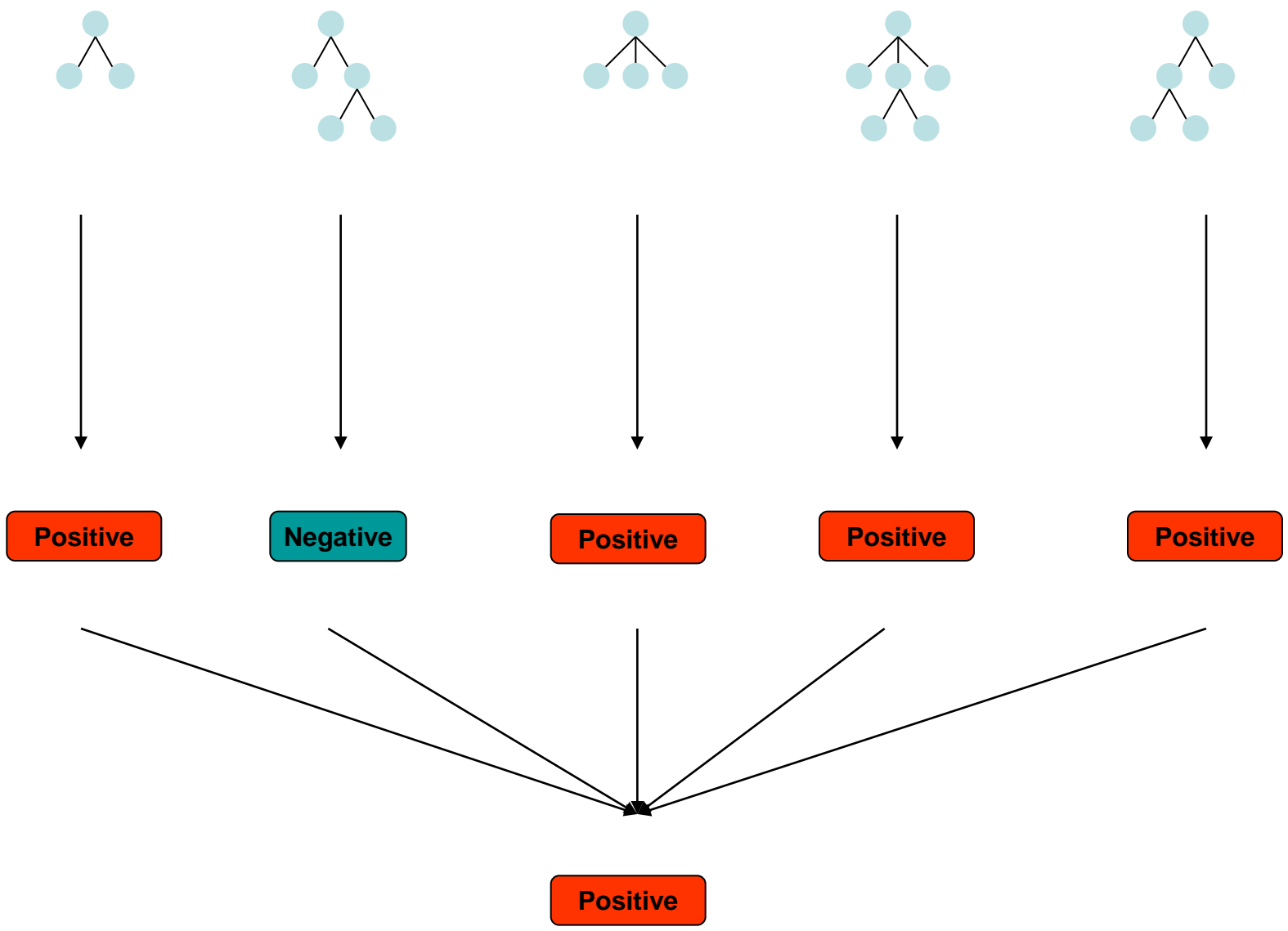
Zhenyu Lu⁺, Xindong Wu⁺, Xingquan
Zhu^{*} and Josh Bongard⁺

⁺University of Vermont

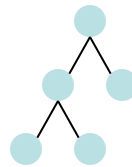
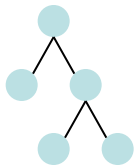
^{*}University of Technology, Sydney



Are all members in an ensemble necessary?[1]



[1]: Margineantu et al., 1997



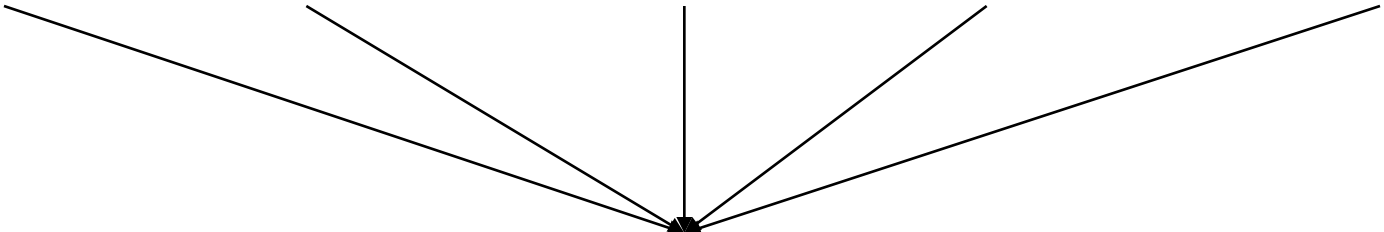
Positive

Negative

~~Positive~~

~~Positive~~

Positive

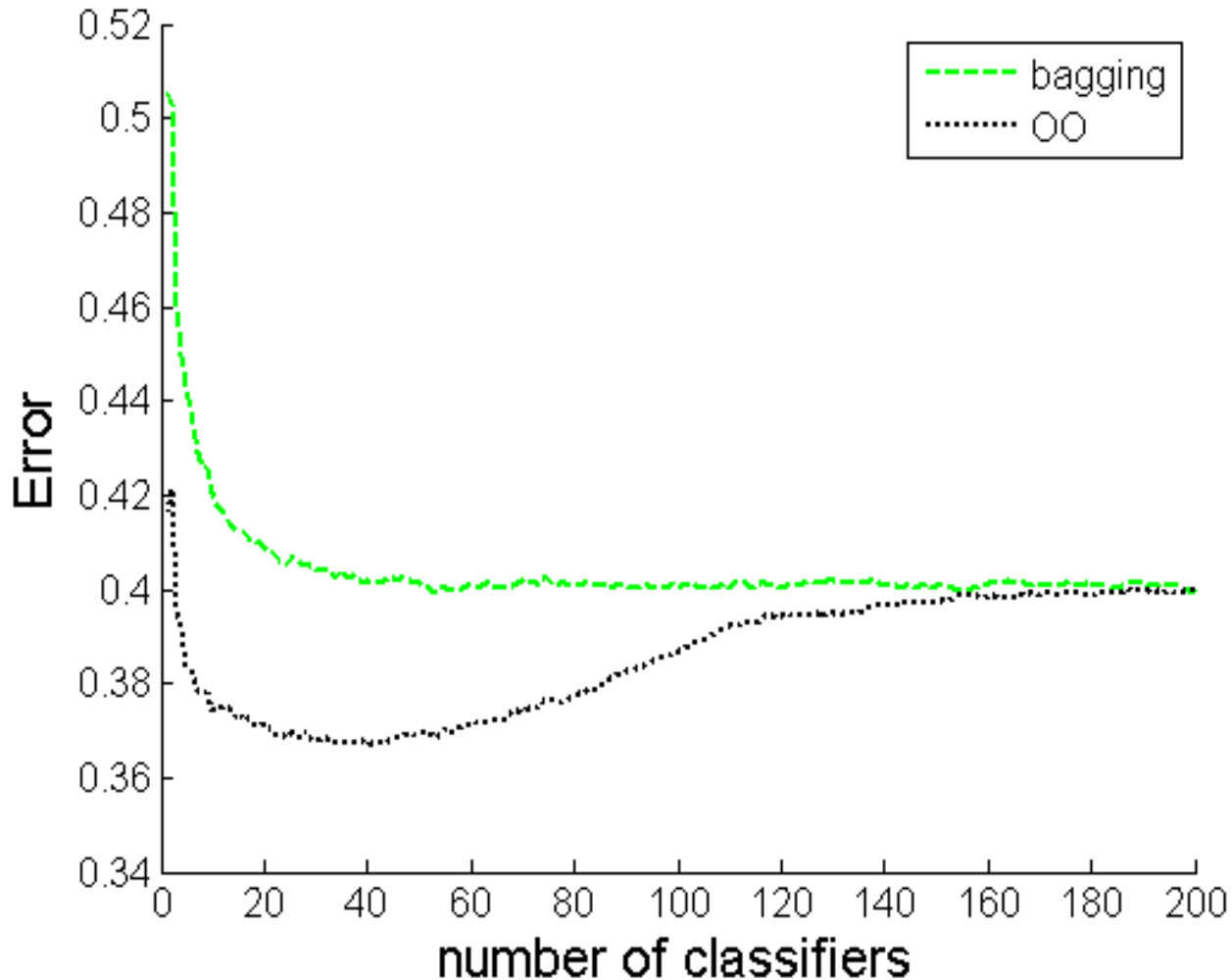


Positive



- Many could be better than all^[2]
- Orientation Ordering^[3] (OO)

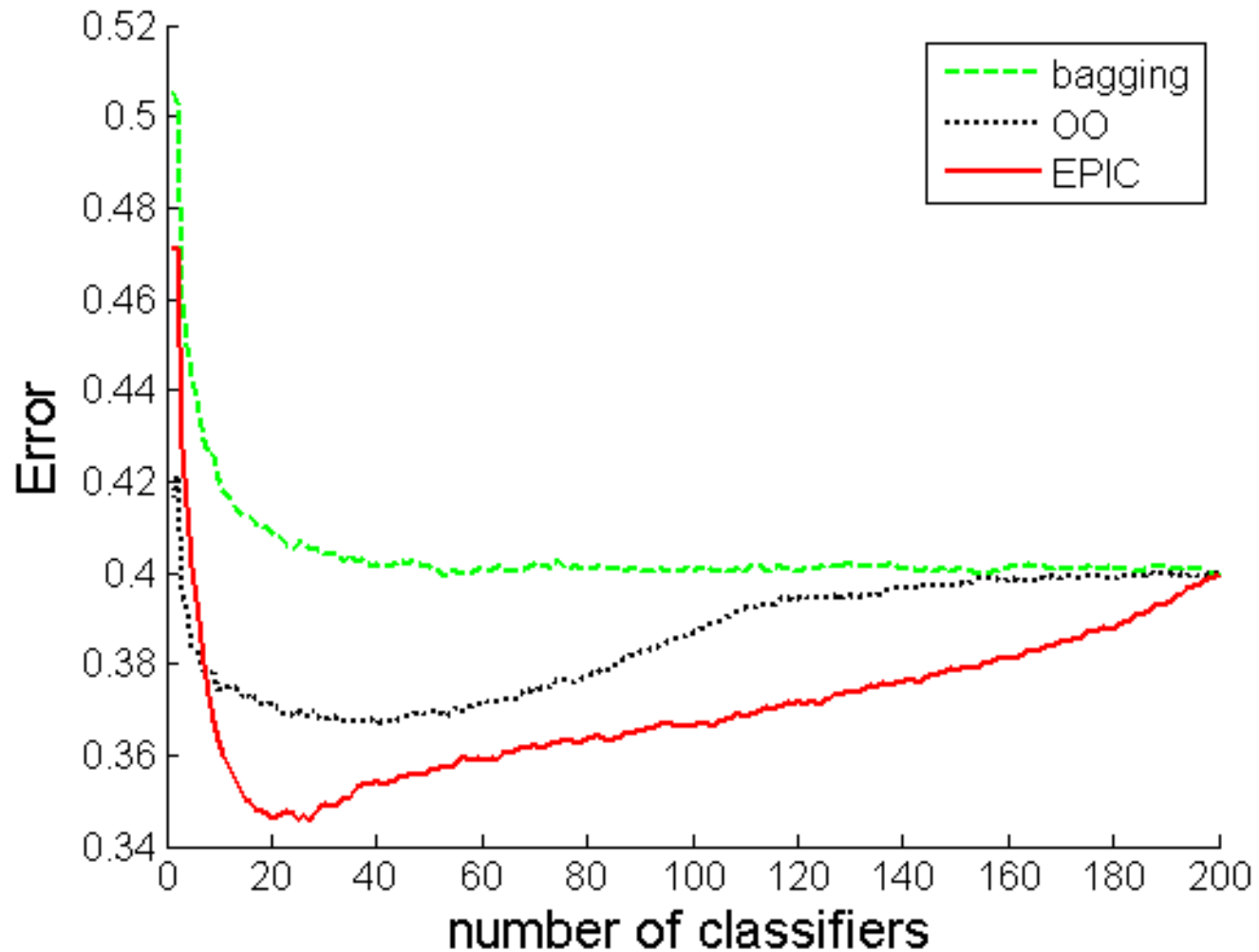
Autos



[2]: Zhou et al., 2002
[3]: Mart'inez et al., 2006

- Ensemble Pruning via Individual Contribution Ordering (EPIC)

Autos





Our Approach

- As for any ensemble method, appropriately handling the accuracy/diversity tradeoff is crucial



Heuristic of Individual Contribution

Correct but in the minority

Correct and in the majority

Positive contribution

Incorrect and in the minority

Incorrect but in the majority

Negative contribution

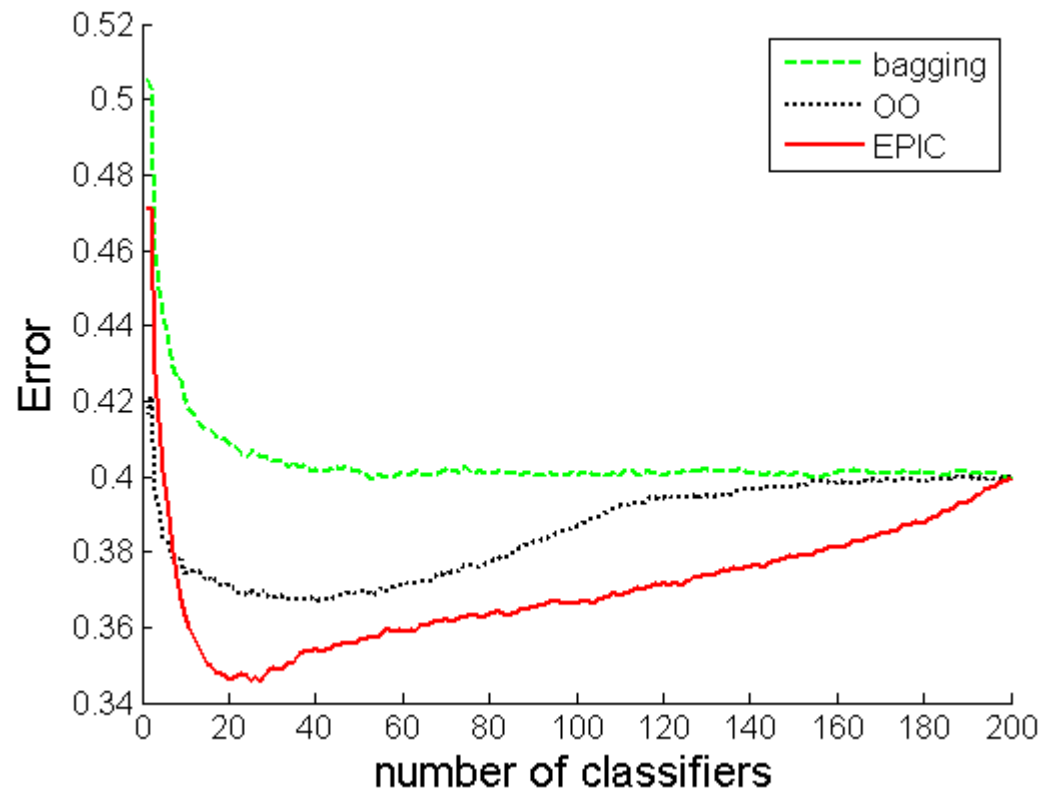


Ensemble Pruning via Individual Contribution ordering: EPIC

- 1) Train an ensemble of classifiers
- 2) Calculate individual contribution of each classifier on a selection set
- 3) Reorder individual classifiers by **decreasing contribution**
- 4) Output the first x percent individual classifiers for prediction; x is given

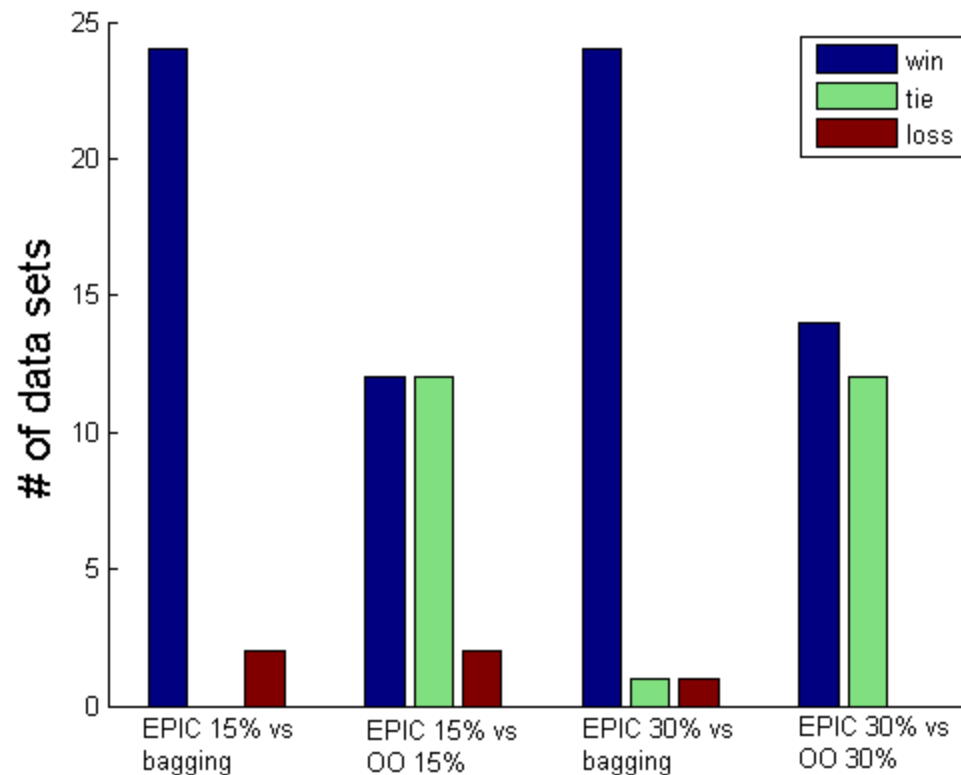
Results – autos

- One data set (autos)
- 300 independent runs, ensemble size 200
- Base learner: J48 decision tree
- ensemble method: bagging



Results – summary

- 26 data set from UCI
- On each set, 300 independent runs, ensemble size 200
- Base learner: J48 decision tree
- ensemble method: bagging



- EPIC outperforms both bagging and OO on two different settings



Conclusions and Future work

- Contribution in decreasing order: correct but in the minority, correct and in the majority, incorrect and in the minority, incorrect but in the majority
- EPIC is a single-parameter, fast (pruning time $O(m \log m)$) and effective (outperforms OO) pruning method
- Future work
 - Make EPIC parameterless
 - Test and generalize EPIC with different base classifier types, different ensemble methods and heterogeneous ensembles
 - Design better individual contribution measure