EPSRC Engineering and Physical Sciences Research Council

ORCHID

Decentralised Independent Bayesian Classifier Combination

Steven Reece, Edwin Simpson and Stephen Roberts

Pattern Analysis and Machine Learning Research Group

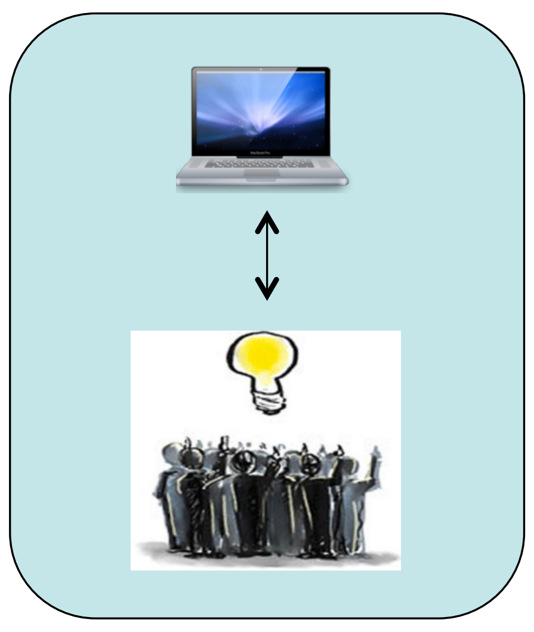
Oxford University

The 'Centralised' IBCC

Combines multiple classifier outputs and accommodates classifier reliability (i.e. trust) in a principled information theoretic way.

IBCC is an unsupervised approach which exploits the latent structure within the data to learn both the reliability of each classifier and the true class labels. This contrasts with traditional supervised approaches which require labelled training data and which are grounded using only this training data. By exploiting the latent structure over all data the IBCC provides a better classifier performance.

Efficient inference can be performed using variational Bayes (VB). However, until recently, only single 'centralised' IBCC platforms have been investigated.



amazonmechanical turk
Artificial Artificial Intelligence

Why Decentralise the IBCC?

Distributed databases

E.g. Multi-platform crowdsourcing with heterogeneous task requirements (one platform classifies images and another documents, for example).

Assume worker accuracy is independent of task type or conditional on type clusters.

Parallel computation

E.g. Topic label newspaper articles from different media groups. These documents are typically archived on newsgroup owned servers (the 'platforms'). Can assume topic label distributions are homogeneous across platforms.

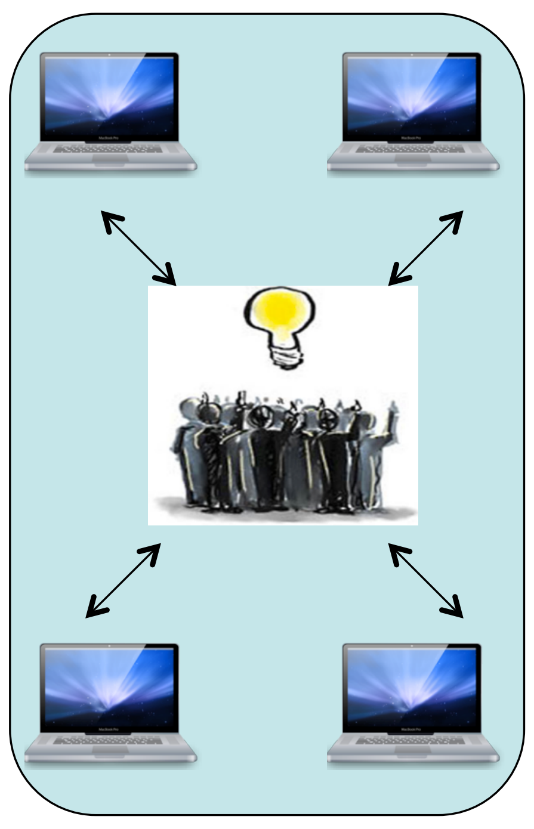
Mitigate the need to transfer big data onto a single server.

Example: Multi-Platform Crowdsourcing

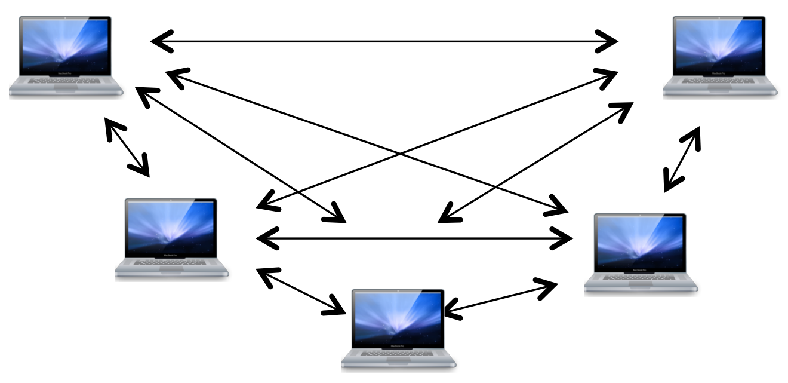
Aim to infer object classes using (possibly) unreliable worker responses and data latent structure when data is **distributed** across platforms and where crowdsourcing applications share a worker pool.

Requirements and consequences:

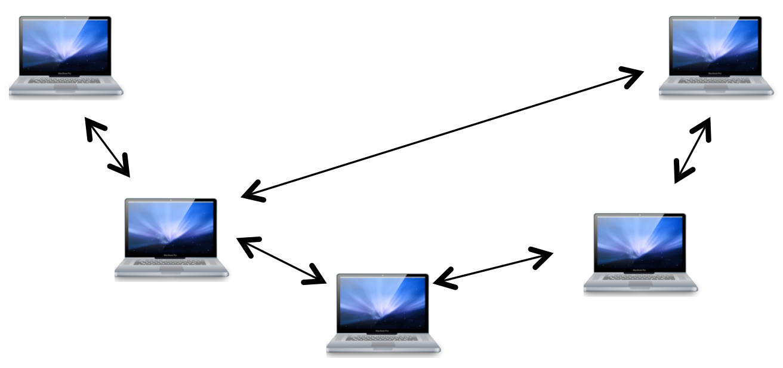
- communication of relevant data between platforms
  Worker accuracy pseudo counts
  Object class pseudo counts
- platforms compete for workers
- coordination of tasks between platforms



Network Infrastructures



Fully connected



Tree or loopy connected

The Decentralised IBCC

Example: Heterogeneous databases (object types differ between platforms)

$t_i^{(p)}$ True label of ith object on platform p.

$c_i^{(pk)}$ Worker 'k' assigned label for ith object on platform p.

$\pi^{(pk)}$ Platform p's confusion matrix for worker k.

$\mu_S^{(k)}$ Worker k's binary accuracy vector (S='+' correct, S='-' wrong).

$\alpha^{(k)}$ Shape parameters of Beta distribution over $\mu^{(k)}$.

$\nu^{(p)}$ Object class Dirichlet distribution parameters.

J(p) Number of object classes on platform p.

$M_{+q}^{(pk)}$ Worker k correct response (+) pseudo count sent by p to q.

$M_{-q}^{(pk)}$ Worker k erroneous response (-) pseudo count sent by p to q.

$$\alpha_S^{(pk)} = \alpha_{0,S}^{(pk)} + N_S^{(pk)} + \sum_{p' \in Neigh(p)} M_{Sp}^{(p'k)} \quad \text{where } S = + \text{ or } S = -.$$

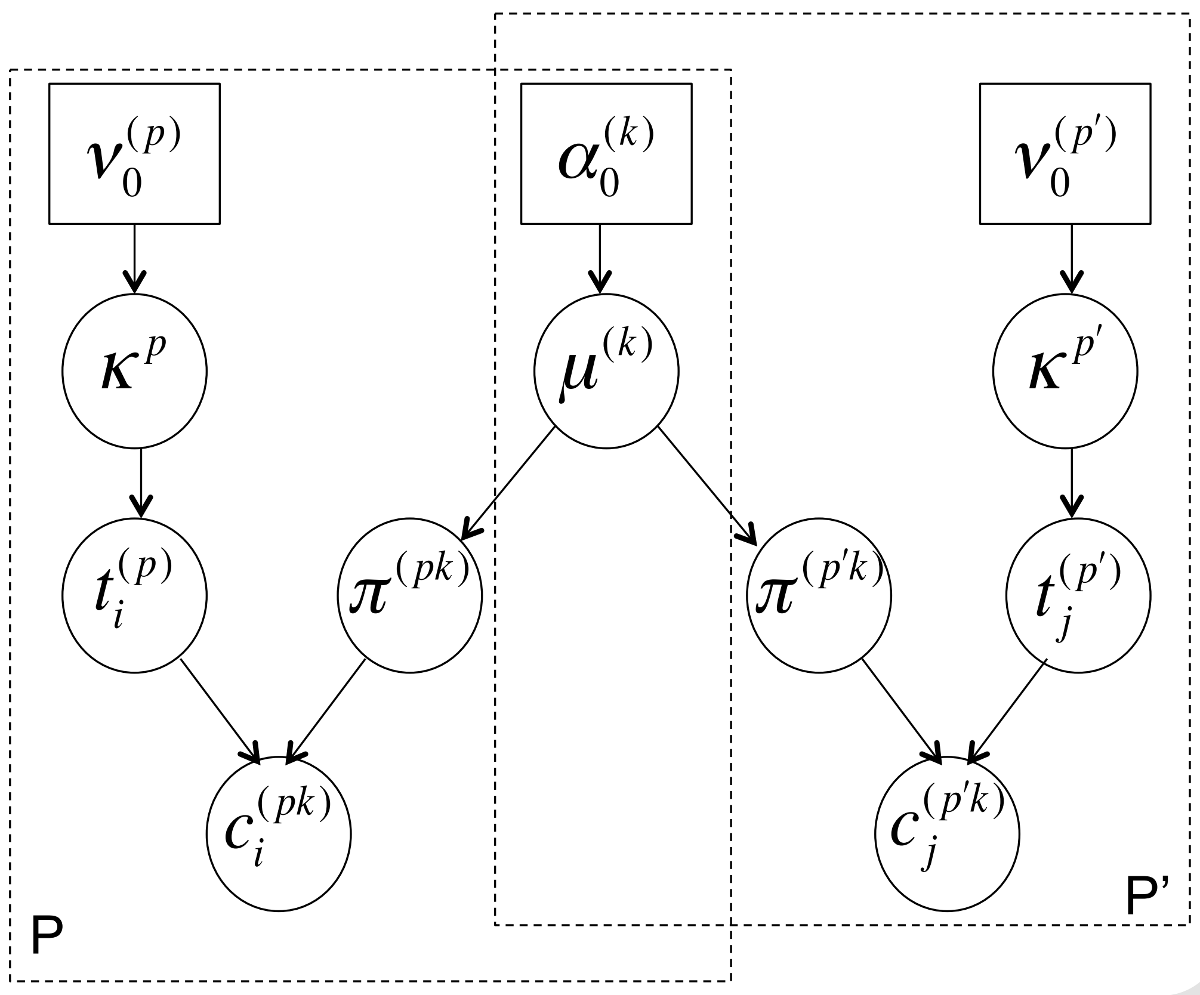
$$M_{Sq}^{(pk)} = N_S^{(pk)} + \sum_{p' \in Neigh(p) \backslash q} M_{Sp}^{(p'k)}$$

$$N_+^{(pk)} = \sum_{j=1}^{J(p)} \sum_{i=1}^{I} [c_i^{(pk)} = j] \mathrm{E}_t[t_i^{(p)} = j]$$

$$N_-^{(pk)} = \sum_{j=1}^{J(p)} \sum_{i=1}^{I} [c_i^{(pk)} \neq j] \mathrm{E}_t[t_i^{(p)} = j]$$

$$\pi_{ii}^{(pk)} = \mu_+^{(pk)} \text{ and } \pi_{ij}^{(pk)} = \frac{\mu_-^{(pk)}}{J(p)-1}$$

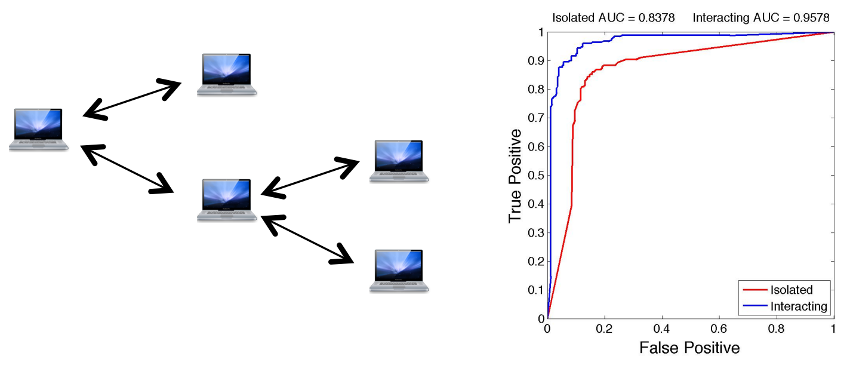
Other VB equations as per centralised IBCC.



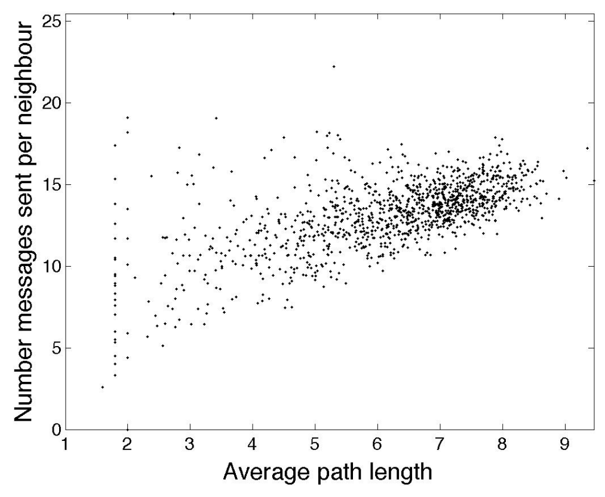
Results: Tree Connected Heterogeneous

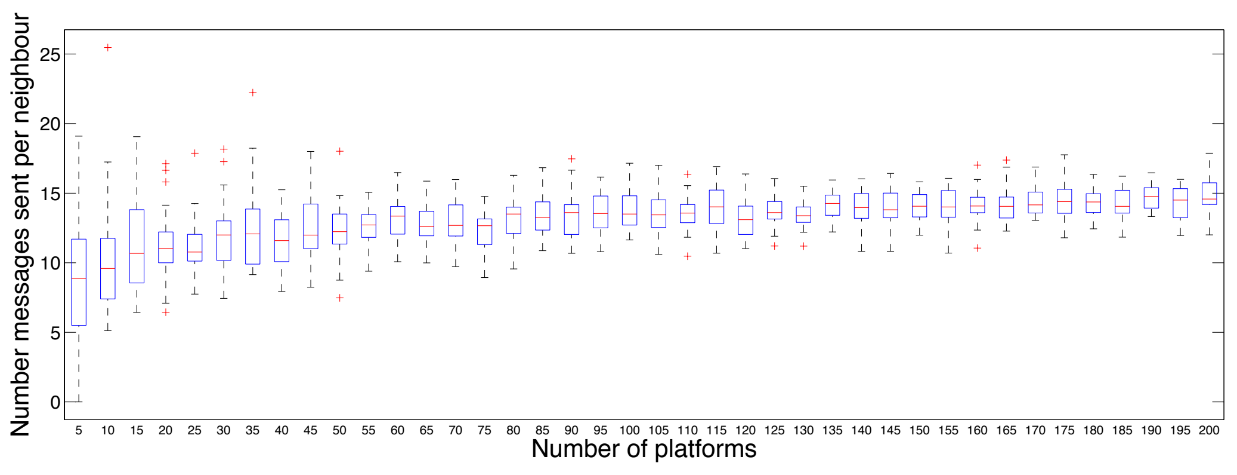
Isolated platforms versus tree connected platform network …

- Between 5 and 200 platforms in network.
- Experiment (simulatation) uses 30 samples for each network size.
- 2 unreliable classifiers (totally random responses).
- 3 reliable classifiers (30% random label error rate).
- 10 objects classified by each platform.
- Each platform has 80% probability of employing each of the workers.
- Platforms communicate new pseudo counts, M, after every VB iteration.



Example network and corresponding average RoC across platforms.





Future Work

Incentive Engineering

Incentivise platforms to communicate worker reliability and class pseudo counts.

Agile Teaming

Platform cliques (fully connected) utilise pseudo counts most efficiently.

Accountable Information Infrastructure

Mitigate message double counting in loopy networks with knowledge of local network structure.

Flexible Autonomy

Platforms identify objects to label or tasks to perform.
Workers choose whether to perform task or not and may propose new tasks.

BAE SYSTEMS

ACFR
AUSTRALIAN CENTRE FOR FIELD ROBOTICS

SECURE

UNIVERSITY OF OXFORD

UNIVERSITY OF Southampton