UNIVERSITY OF Southampton

Bayesian Aggregation of Categorical Distributions with Applications in Crowdsourcing

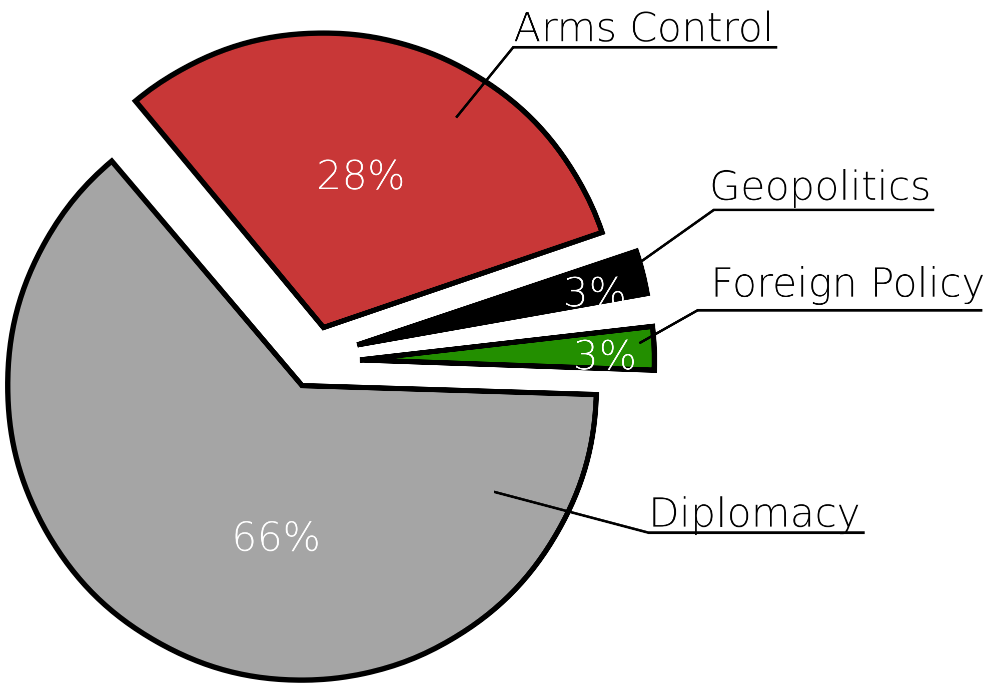
A. Augustin[1], M. Venanzi[2], A. Rogers[3], N.R. Jennings[4]

[1] Southampton University, aa7e14@ecs.soton.ac.uk
[2] Microsoft London, mavena@microsoft.com
[3] Oxford University, alex.rogers@cs.ox.ac.uk
[4] Imperial College, n.jennings@imperial.ac.uk

Motivating Examples

Possible application domain include information retrieval and recommendation:

Source: BBC, 2017



Example judgment of proportions

Challenges

- Malicious participants (i.e. **spammers**) constitute up to 45% of all workers [Vuurens et al., 2011].

- Spammers increase the cost of judgment acquisition and degrade accuracy of the aggregation.

Previous Approaches

- Linear opinion pool (arithmetic mean)
  - ✔ Fast
  - ✖ Does not account for spammers
- Label distribution learning (LDL) framework [Geng, 2016]
  - ✔ Comprehensive
  - ✖ Does not account for spammers
- Independent Bayesian classifier combination (IBCC) [Kim et al, 2012]
  - ✔ Accounts for spammers
  - ✖ Does not aggregate distributions

Our Solution (MBCC)

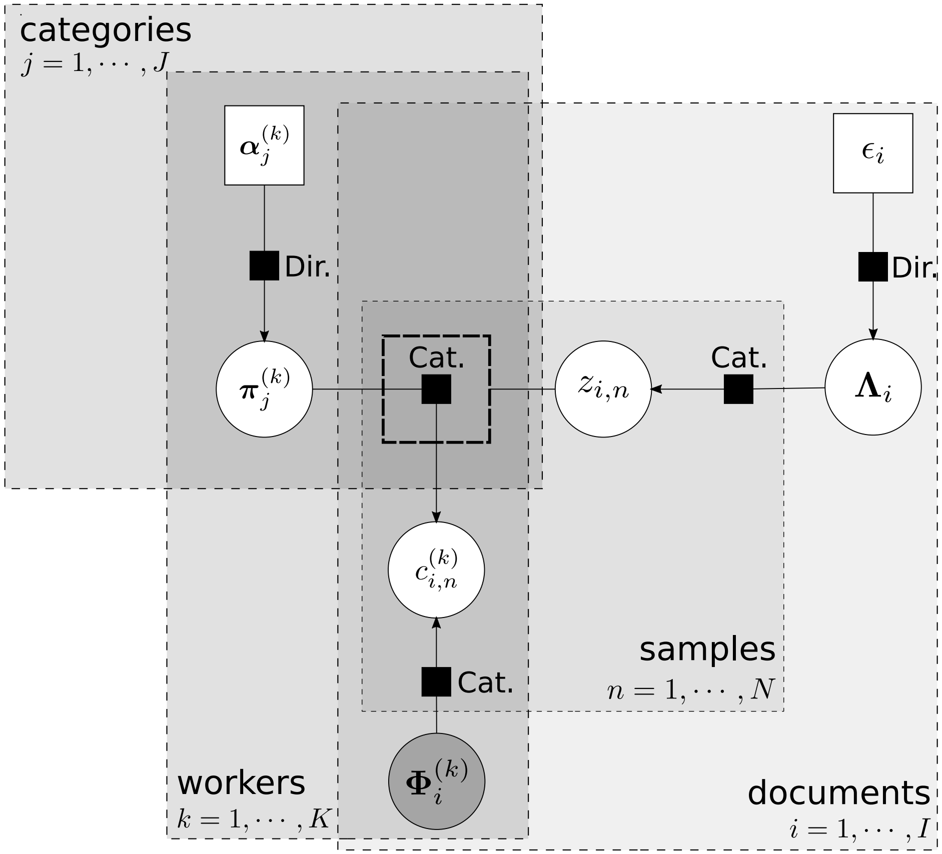
$\boldsymbol{\pi}_j^{(k)} \sim \mathrm{Dir}\left(\boldsymbol{\alpha}_j^{(k)}\right)$ j-th row of the confusion matrix of worker *k*

$\boldsymbol{\Lambda}_i \sim \mathrm{Dir}\left(\epsilon_i\right)$ Aggregated distribution of document *i*

$\mathrm{z}_{i,n} \sim \mathrm{Cat}\left(\boldsymbol{\Lambda}_i\right)$ Sampled category *n* from aggregated distribution of document *i*

$\mathrm{c}_{i,n}^{(k)} \sim \mathrm{Cat}\left(\boldsymbol{\pi}_{z_{i,n}}^{(k)}\right)$ Sampled category *n* from worker *k*'s judgment of document *i*

$\Phi_i^{(k)}$ Judgment of document *i* by worker *k* (categorical dist.)



Factor graph of MBCC used for inference

Datasets

1. **SemEval**: 6 sentiments in 100 news headlines (1,000 judgments)
2. **IAPR-TC12**: 6 regions in 16 urban and rural scenes (441 judgments)
3. **Colours**: 10 colours in 20 countries' flag (460 judgments)

Task 55: Making peace from victory over poverty.

Joy [70]	Sadness [0]	Disgust [35]
Anger [0]	Surprise [0]	Fear [0]

SemEval

IAPR-TC12

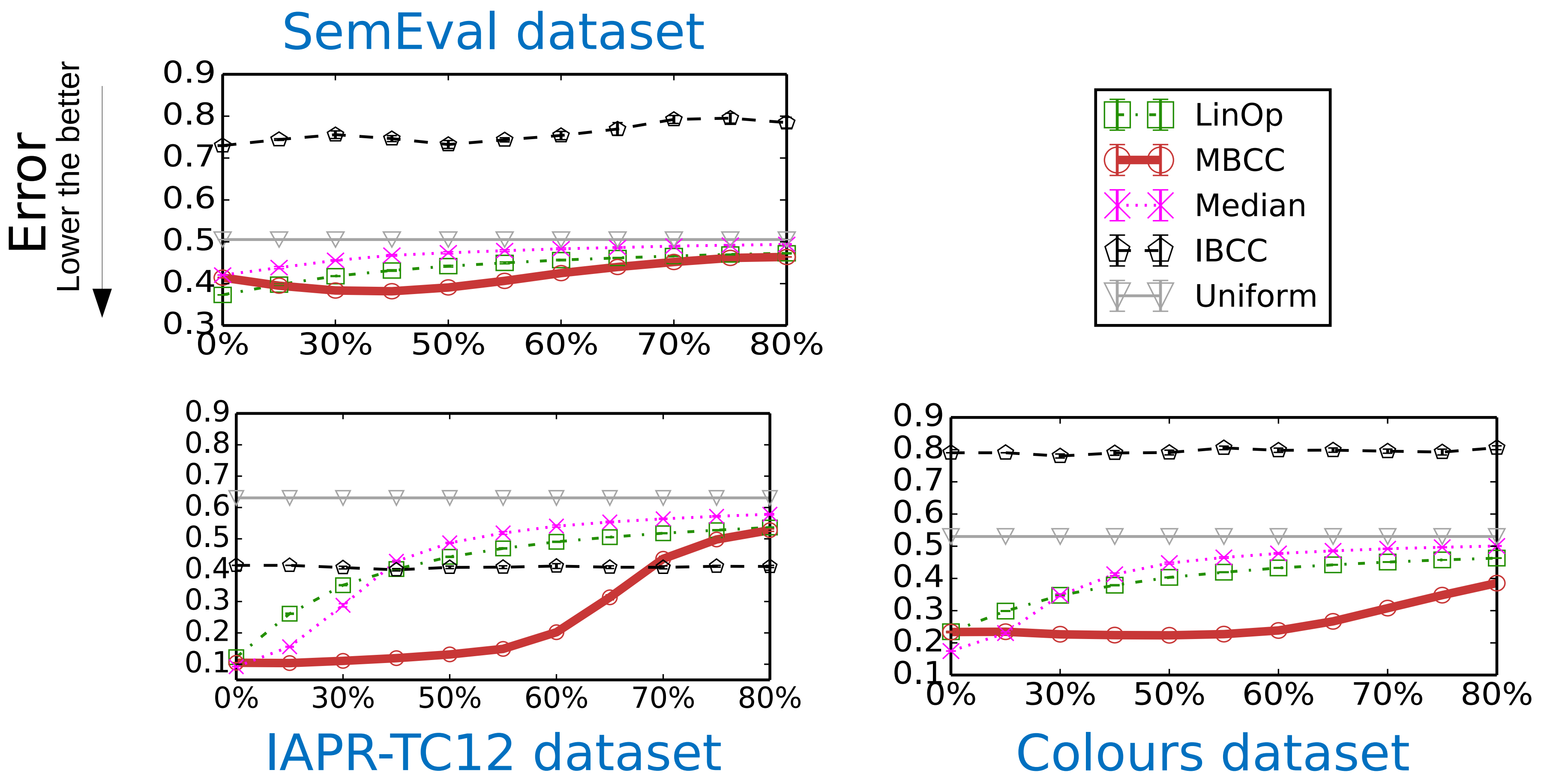
Colours

- Each dataset is augmented with additional synthetic spammers to explore the loss of accuracy as they increase in number.
- The distribution of each spammer *k* for each document *i* shares a prior Dirichlet distribution such that

$$\Phi_i^{(k)} \sim \mathrm{Dir}\left(\mathbf{1}\right).$$

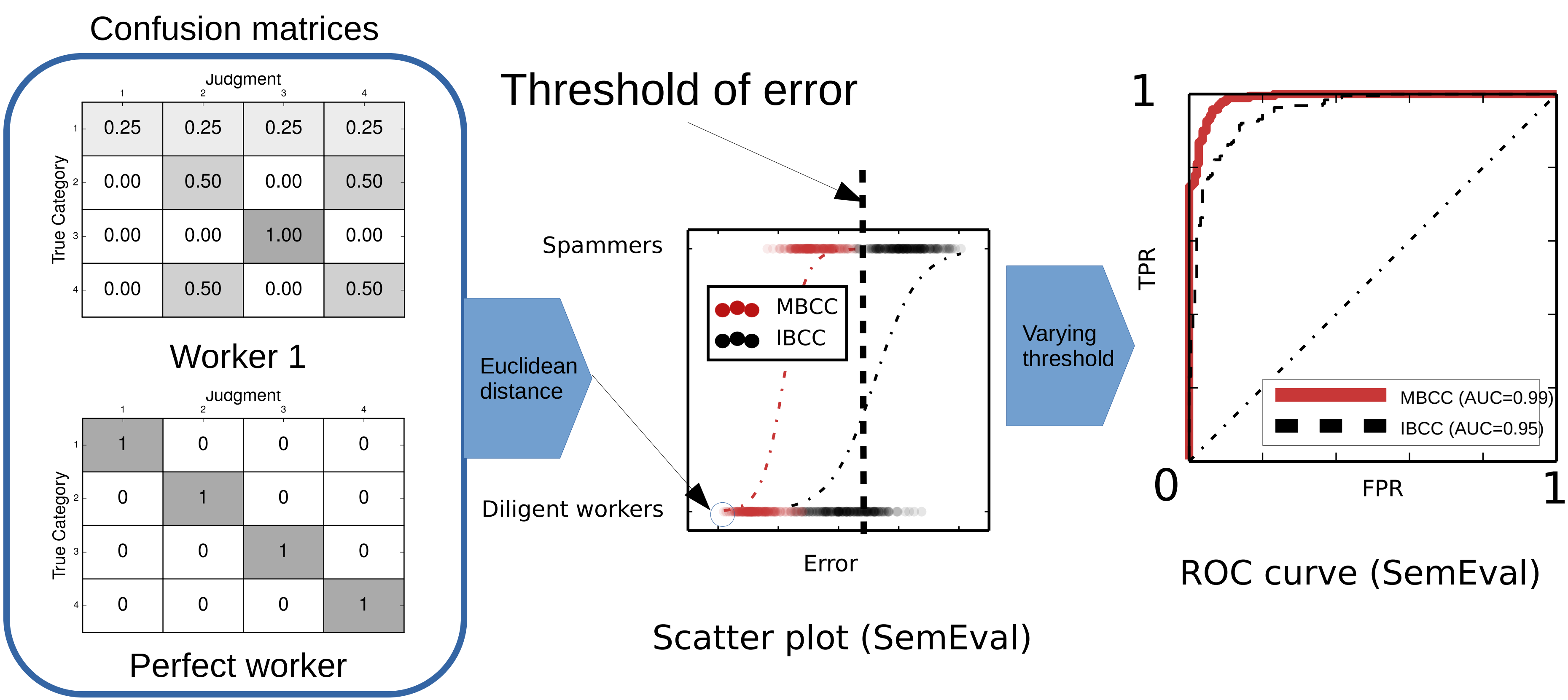
Results 1: Aggregation Error

Aggregation error is robust to an increasing ratio of spammers:



Results 2: Classification of Workers

Accuracy of classifying diligent workers from spammers is improved:



References

[1] Vuurens et al., "How Much Spam Can You Take? An Analysis of Crowd sourcing Results to Increase Accuracy", CIR, 2011.

[2] X. Geng, “Label distribution learning”. IEEE Transactions on Knowledge and Data Engineering, 2016.

[3] Kim et al., "Bayesian Classifier Combination", AISTATS, 2012.

[4] M. Bacharach, “Normal Bayesian Dialogues”, Journal of the American Statistical Association, 1979.