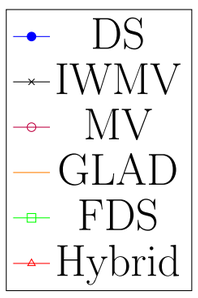
Fast Dawid-Skene : Supplementary Results

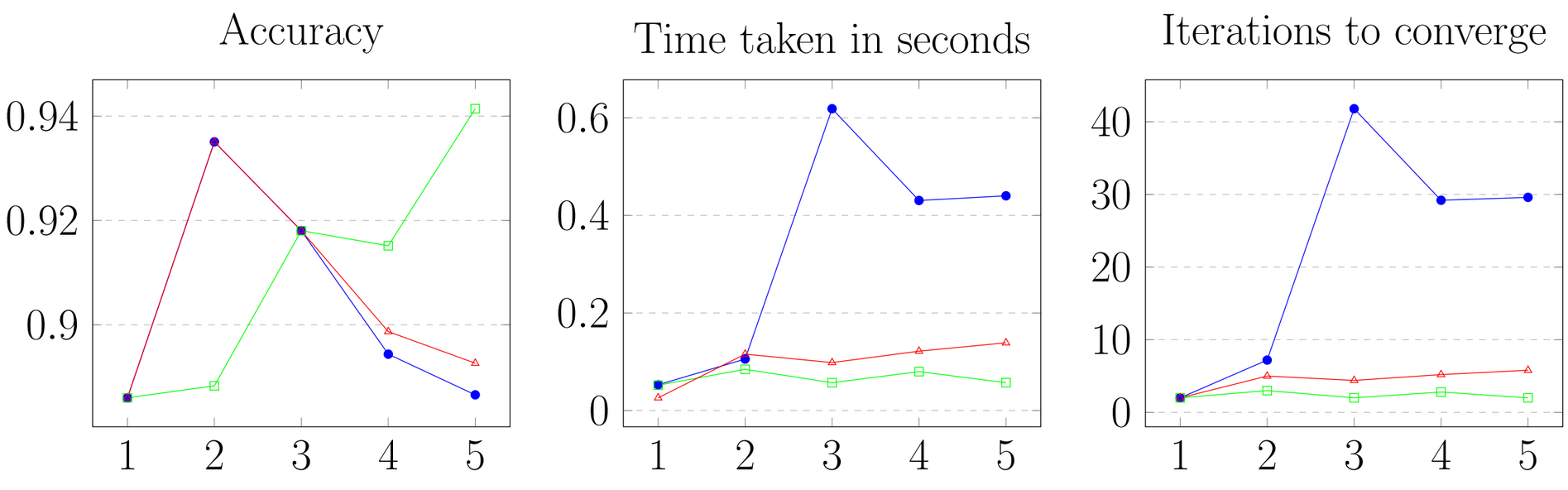
Contents

The legend for all the plots in this document is as follows:



1 Multiple Correct Options

The following are the results obtained in the multiple correct answer setting on the *AffectAnnotation* dataset. These correspond to averaging over five subsets of annotators. The results are reported for a varying number of annotators from one to five, which is shown on the x-axis.



2 Online Vote Aggregation

This section contains results for the extension of the FDS to the online vote aggregation setting. Results comparing the online versions of DS, FDS, and Hybrid are provided for multiple datasets at their respective maximum number of annotators. Results comparing regular FDS and online FDS are also shown for a varying number of annotators for each dataset. For each dataset, the batch size, which is the initial number of data points that were aggregated using the regular version of the respective algorithms, before switching to online, is also specified. This batch size varied with the size of the full dataset.

Adult2

Number of annotators = 9
Batch Size = 100

	DS	FDS	Hybrid
Accuracy	77.38%	75.41%	75.41%
Time taken to converge (s)	0.51	0.31	0.43
# Iterations to converge	15	2	8

Table 1: Online Vote Aggregation on *Adult2* dataset

Accuracy	2	3	4	5	6	7	8	9
FDS	69.51%	74.10%	77.05%	74.43%	75.08%	75.41%	75.41%	75.74%
Online FDS	69.84%	75.41%	77.05%	74.43%	75.41%	75.41%	75.74%	75.41%

Table 2: Online FDS vs FDS for varying number of annotators.

BM

Number of annotators = 5
Batch Size = 250

	DS	FDS	Hybrid
Accuracy	70.40%	69.60%	69.80%
Time taken to converge (s)	1.18	0.45	0.49
# Iterations to converge	94	3	6

Table 3: Online Vote Aggregation on *BM* dataset

Accuracy	2	3	4	5
FDS	67.50%	67.90%	68.80%	69.70%
Online FDS	67.20%	68.80%	68.60%	69.60%

Table 4: Online FDS vs FDS for varying number of annotators.

TREC2010

Number of annotators = 5
Batch Size = 1000

	DS	FDS	Hybrid
Accuracy	54.34%	54.23%	55.24%
Time taken to converge (s)	25.53	10.45	12.27
# Iterations to converge	67	8	15

Table 5: Online Vote Aggregation on *TREC2010* dataset

Accuracy	2	3	4	5
FDS	43.73%	49.95%	53.74%	55.40%
Online FDS	42.97%	48.12%	51.58%	54.23%

Table 6: Online FDS vs FDS for varying number of annotators.

VideoAffect

Number of annotators = 10
Batch Size = 1000

	DS	FDS	Hybrid
Accuracy	52.44%	48.06%	47.19%
Time taken to converge (s)	46.04	23.39	27.73
# Iterations to converge	45	4	10

Table 7: Online Vote Aggregation on *VideoAffect* dataset

Accuracy	2	3	4	5	6	7	8	9	10
FDS	40.34%	44.32%	48.08%	48.57%	50.52%	52.38%	51.94%	51.17%	49.85%
Online FDS	42.11%	42.87%	46.95%	47.97%	48.12%	50.17%	49.89%	49.09%	48.06%

Table 8: Online FDS vs FDS for varying number of annotators.

LabelMe

Number of annotators = 3
Batch Size = 100

	DS	FDS	Hybrid
Accuracy	76.40%	76.23%	77.93%
Time taken to converge (s)	1.44	0.92	1.08
# Iterations to converge	25	2	8

Table 9: Online Vote Aggregation on *LabelMe* dataset

Accuracy	2	3
FDS	74.87%	77.08%
Online FDS	72.67%	76.23%

Table 10: Online FDS vs FDS for varying number of annotators.

RTE

Number of annotators = 10
Batch Size = 200

	DS	FDS	Hybrid
Accuracy	90.38%	91.25%	91.75%
Time taken to converge (s)	0.65	0.55	0.61
# Iterations to converge	20	5	11

Table 11: Online Vote Aggregation on *RTE* dataset

Accuracy	2	3	4	5	6	7	8	9	10
FDS	81.25%	87.25%	89.75%	91.50%	92.25%	91.25%	92.38%	92.50%	91.88%
Online FDS	79.12%	87.25%	89.38%	90.88%	91.38%	90.88%	91.12%	91.38%	91.25%

Table 12: Online FDS vs FDS for varying number of annotators.

SP

Number of annotators = 5
Batch Size = 1000

	DS	FDS	Hybrid
Accuracy	90.94%	90.60%	90.64%
Time taken to converge (s)	4.40	3.76	4.09
# Iterations to converge	26	4	5

Table 13: Online Vote Aggregation on *SP* dataset

Accuracy	2	3	4	5
FDS	85.59%	88.41%	90.02%	90.74%
Online FDS	83.57%	88.06%	89.90%	90.60%

Table 14: Online FDS vs FDS for varying number of annotators.

3 Results of FDS and Hybrid

This section contains experimental results on accuracy, time taken, and iterations taken with larger sized graphs. The x-axis denotes the number of annotators. For each dataset, the number of annotators was varied, by choosing a fixed number of annotations randomly at each step.

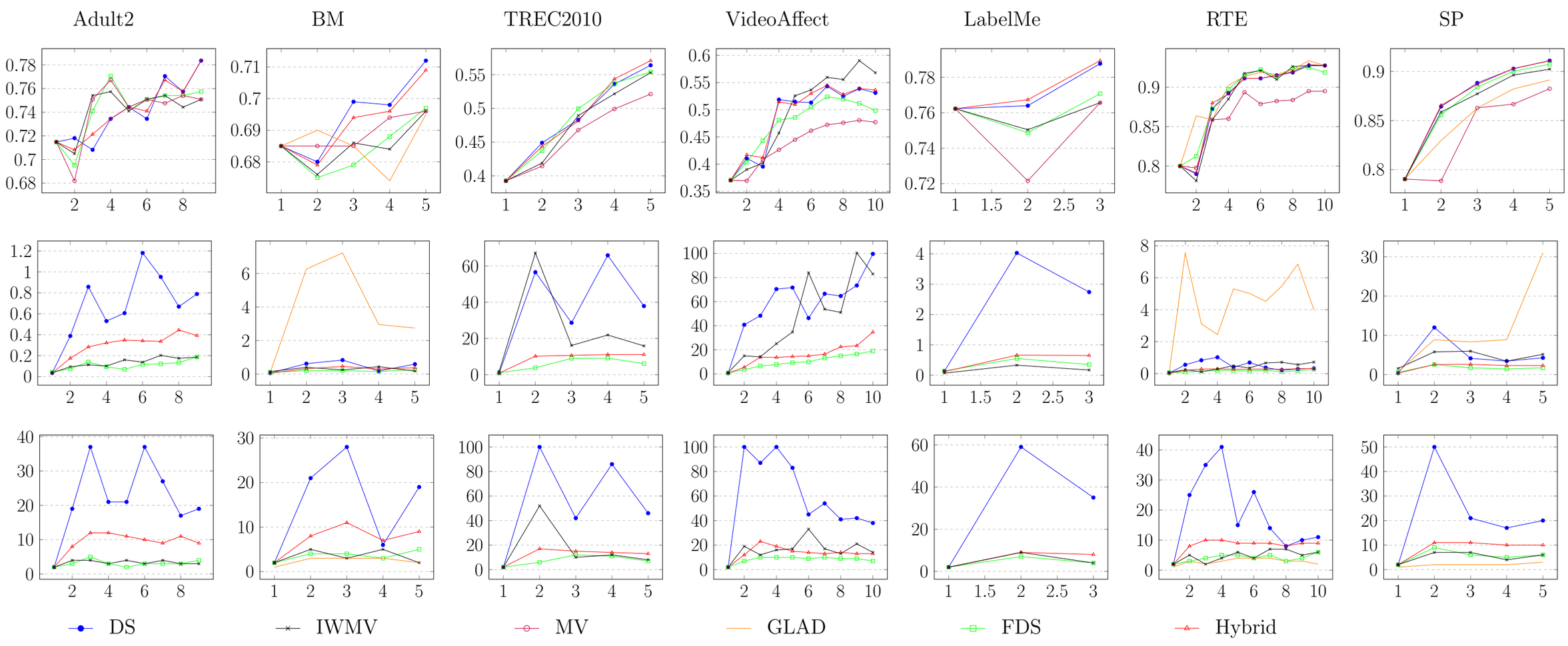


Figure 1: Experimental results: *(Row 1:)* Accuracy of different methods across the considered datasets; *(Row 2:)* Time taken in seconds to converge; and *(Row 3:)* Number of iterations to converge. X-axis denotes the varying number of annotators studied for each dataset.