Introduction

- Crowdsourcing provides huge opportunities and scalability solutions for grading large scale tasks, such as MOOCs.
- Reliability and quality of graders and crowdsourced data are challenging issues.
- Workers might give random grades, which are spam; or provide biased grades, which need to be corrected.
- The budget for hiring graders is limited, in many cases.

Grading through Crowdsourcing Applications

- Grading large scale classes (MOOCs)

Thousands of students submissions

- Labeling kid-friendly images

No adult content?

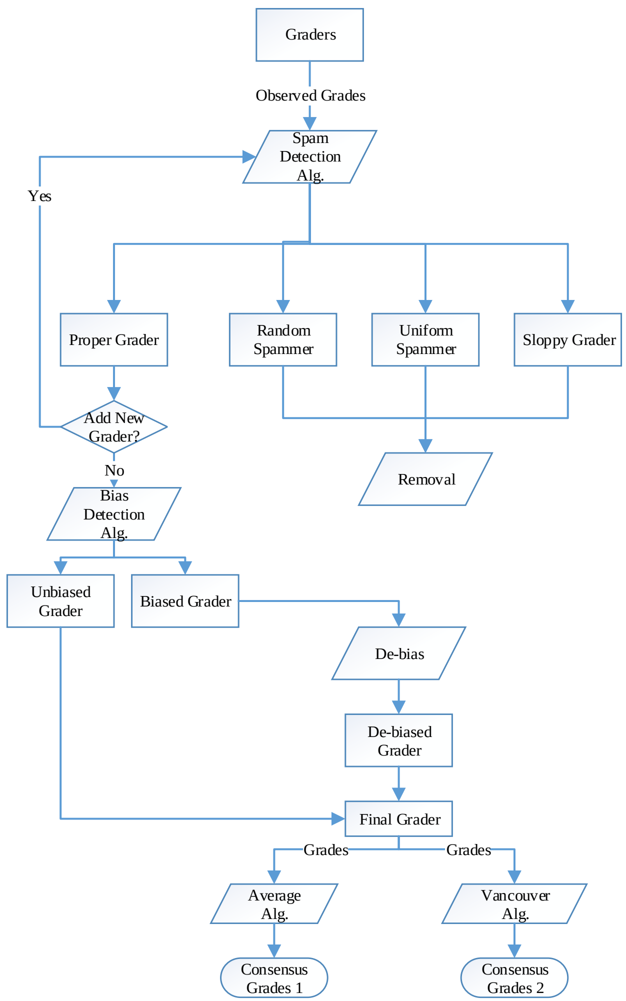
Content requires parental guidance?

Mainly for adults …

Research Purpose

- Examine the influence of the spammers on grading complex tasks
- Build a crowdsourcing framework to combine spam detection and de-biasing algorithms to optimize the estimated true grades
- Analyze impact of the graders’ number on the estimated true grades
- Optimize the cost by reducing the number of graders

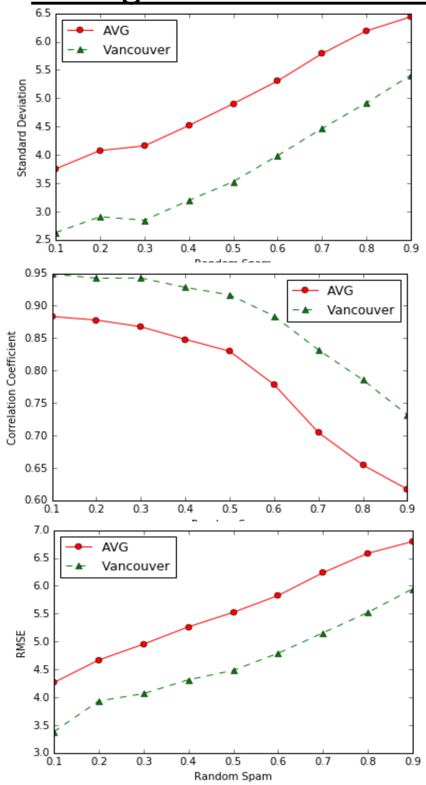
Methodology



Experimental Results

- Evaluation Metrics – standard deviation (σ); coefficient correlation (ρ) ; RMSE

<u>Each grader review 6 tasks</u>



Impact of spam proportion on estimated true grades

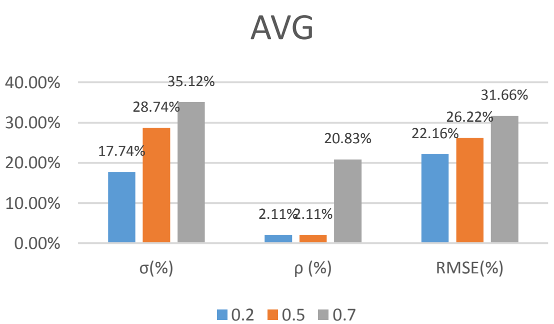
Apply framework

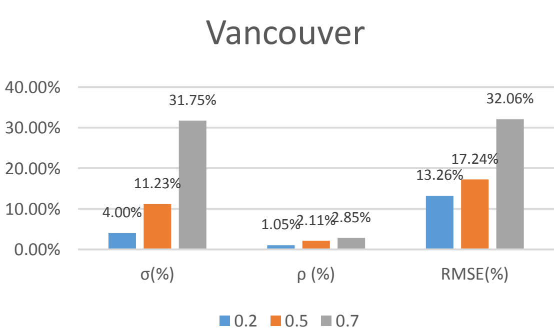
Compare without spam removing or de-biasing

Metrics			σ	ρ	RMSE
Subm_grades =4	AVG	Spam	6.13	0.77	6.86
		Spam Filter	3.79	0.95	3.50
		Spam Filter+Debias	2.80	0.96	2.93
	Vancouver	Spam	5.63	0.83	5.82
		Spam Filter	3.97	0.92	4.20
		Spam Filter+Debias	3.09	0.95	3.23
Subm_grades =6	AVG	Spam	4.96	0.85	6.02
		Spam Filter	2.93	0.95	3.34
		Spam Filter+Debias	2.41	0.97	2.60
	Vancouver	Spam	4.12	0.90	4.90
		Spam Filter	3.03	0.95	3.77
		Spam Filter+Debias	2.91	0.96	3.20
Subm_grades =10	AVG	Spam	4.13	0.91	4.79
		Spam Filter	2.10	0.97	2.96
		Spam Filter+Debias	1.88	0.98	2.43
	Vancouver	Spam	2.55	0.96	3.60
		Spam Filter	2.23	0.97	3.01
		Spam Filter+Debias	1.96	0.97	2.89

Experimental Results

<u>Impact of different ratios of biased graders</u>





<u>Impact of different rations of spammers</u>

Subm_grades = 6			σ	ρ	RMSE
Rand = 0.1 Uniform = 0.1 Sloppy = 0.1	AVG	Spam	4.96	0.85	6.02
		Spam Filter	2.93	0.95	3.34
	Vancouver	Spam	4.12	0.90	4.90
		Spam Filter	3.03	0.95	3.77
Rand = 0.4 Uniform = 0.3 Sloppy = 0.2	AVG	Spam	8.06	0.46	8.16
		Spam Filter	2.88	0.95	3.19
	Vancouver	Spam	7.63	0.59	7.86
		Spam Filter	3.51	0.93	4.11
Rand = 0.3 Uniform = 0.2 Sloppy = 0.4	AVG	Spam	6.72	0.66	6.73
		Spam Filter	2.13	0.97	2.94
	Vancouver	Spam	5.47	0.73	5.94
		Spam Filter	2.60	0.96	3.27

<u>Different num. of new graders added</u>

					σ	ρ	RMSE
Rand = 0.1, Uniform = 0.1, Sloppy = 0.1, Bias = 0.2	Subm_ grades = 4	AVG	Spam Filter	n_{thr} = 3	2.71	0.95	2.95
				Full	2.48	0.96	2.58
	Subm_ grades = 6	AVG	Spam Filter	n_{thr} = 3	2.02	0.97	2.41
				Full	1.94	0.98	2.03
	Subm_ grades = 10	AVG	Spam Filter	n_{thr} = 3	2.01	0.97	2.39
				Full	1.81	0.98	2.02
Rand = 0.4, Uniform = 0.3, Sloppy = 0.2, Bias = 0.2	Subm_ grades = 4	AVG	Spam Filter	n_{thr} = 3	2.63	0.96	2.99
				Full	2.52	0.96	2.74
	Subm_ grades = 6	AVG	Spam Filter	n_{thr} = 3	2.58	0.96	2.37
				Full	2.11	0.97	2.12
	Subm_ grades = 10	AVG	Spam Filter	n_{thr} = 3	2.47	0.96	2.26
				Full	2.13	0.97	2.15

Conclusion

- With the framework, we are able to obtain significant improvement up to 32%.
- Fewer graders could be used to get estimated true grades without significant difference compared to original settings for the number of graders.