Sentiment Analysis via Deep Hybrid Textual-Crowd Learning Model

Kamran Ghasedi Dizaji, Heng Huang

kag221@pitt.edu, heng.huang@pitt.edu

Electrical and Computer Engineering Department, University of Pittsburgh, USA

Motivations

- Efficient mining of *public opinions* is very valuable for various industries and businesses.
- *Crowdsourcing* provides a useful platform to employ human skills in *sentiment analysis*.
- Crowdsourcing aggregation models are *incompetent* when the number of crowd labels per worker is *not sufficient* to train parameters, or when it is *not feasible* to collect labels for each sample in a large dataset.
- Crowdsourcing aggregation models do not utilize *text data*, and consider *crowd labels* as the only source of information.

Contributions

- Proposing a *hybrid crowd-text model* for sentiment analysis, consisting of a *generative crowd aggregation model* and a *deep sentimental autoencoder*.
- Defining a *unified* objective function for the hybrid model, and deriving an efficient optimization algorithm to solve the problem.
- Achieving superior or competitive results compared to alternative models, especially when the crowd labels are *scarce*.

(a) *MV-DeepAE* (b) *CrowdDeepAE*

Figure 1: 2D visualization of *CrowdDeepAE* (ours) and *MV-DeepAE* features

CrowdDeepAE Model

Objective function:

$$\max_{\boldsymbol{\theta},\mathbf{W},\mathbf{1}^T\boldsymbol{\alpha}=M+1,\boldsymbol{\alpha}\geq 0} \sum_{ijck} q_{ic}^{(t)} \log\left([d_i]^{\lambda_d}[e_{ic}]^{\alpha_0}[p_{ijck}]^{\alpha_j \mathbf{1}_{ijk}}\right)$$

$$where\; q_{ic}^{(t)} \propto \prod_{jk}(e_{ic})^{\alpha_0}(p_{ijck})^{\alpha_j \mathbf{1}_{ijk}}$$

Algorithm 1: CrowdDeepAE Algorithm

Initialize $\mathbf{q}_i$ by majority voting $\forall i \in \{1, ..., N\}$

while *not converged* **do**

$\quad \min_{\boldsymbol{\theta}} - \sum_{ijck} q_{ic}^{(t)} \log\left([p_{ijck}]^{\alpha_j \mathbf{1}_{ijk}}\right) + \lambda_\theta \sum_j \|\boldsymbol{\theta}_j\|_F$

$\quad \min_{\mathbf{1}^T\boldsymbol{\alpha}=M+1,\boldsymbol{\alpha}\geq 0} \lambda_\alpha \boldsymbol{\alpha}^T\boldsymbol{\alpha} - \boldsymbol{\alpha}^T\boldsymbol{\beta}$

$\quad \min_{W} - \sum_{ic} q_{ic}^{(t)} \log P_W(Y_i = c|\mathbf{X}_i^{Te}) - \frac{\lambda_d}{\alpha_0} \log P_W(\mathbf{X}_i^{Te}|\tilde{\mathbf{X}}_i^D)$

$\quad q_{ic} \propto \prod_{jk}(e_{ic})^{\alpha_0}(p_{ijck})^{\alpha_j \mathbf{1}_{ijk}}$

end

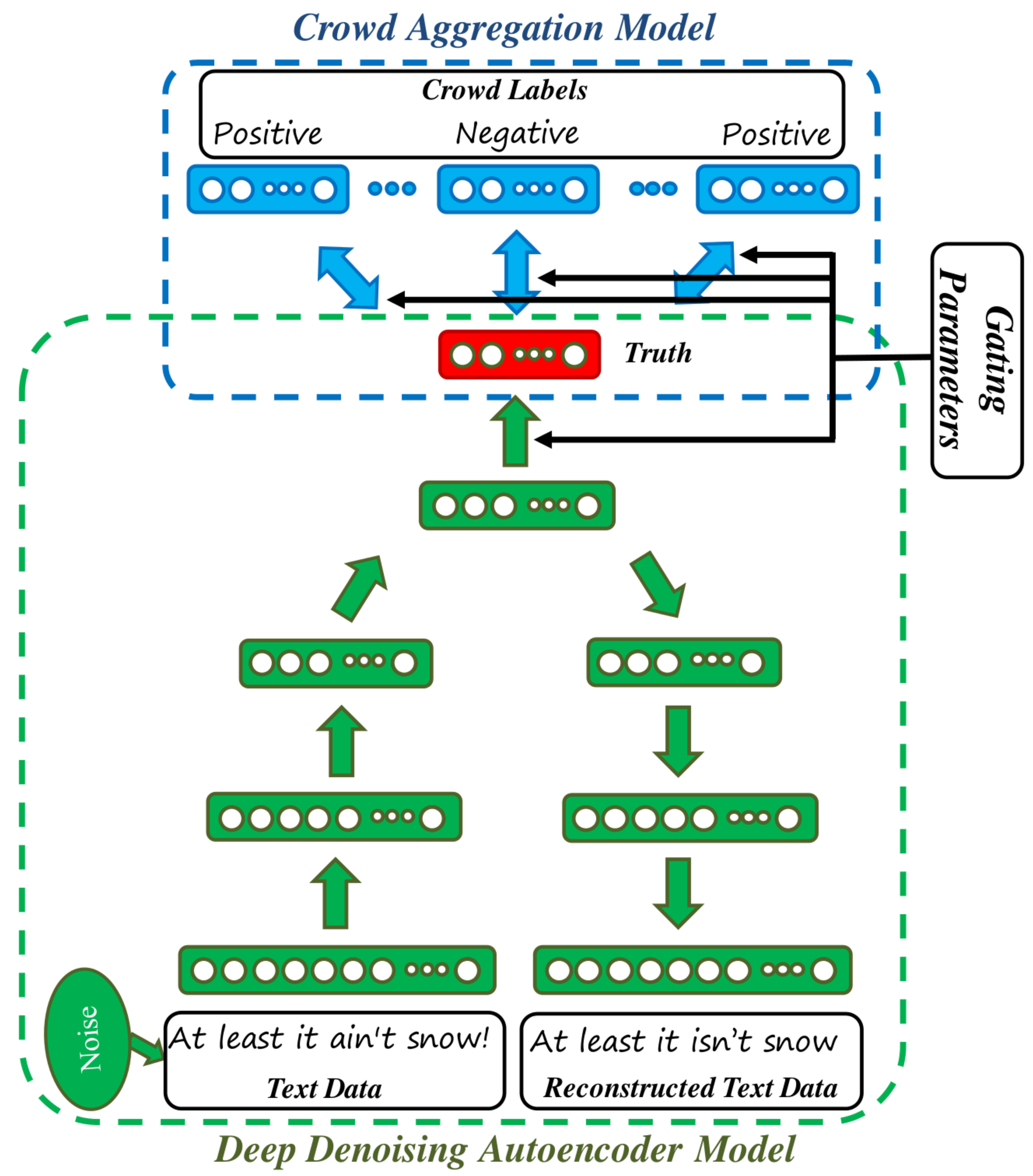


Figure 2: *CrowdDeepAE* architecture.

	Model	*CF* (20% labels) Accuracy	Ave. recall	NLPD	AUC	SP (20% labels) Accuracy	Ave. recall	NLPD	AUC
Crowd	*MV*	0.625	0.550	1.392	0.725	0.710	0.710	1.192	0.704
	IWMV	0.630	0.562	1.368	0.735	0.710	0.710	1.167	0.715
	VD	0.650	0.585	1.252	0.745	0.710	0.710	1.112	0.728
	DS	0.610	0.488	1.285	0.681	0.500	0.500	0.695	0.500
	IBCC	0.688	0.545	0.972	0.822	0.740	0.740	0.516	0.835
	CBCC	0.635	0.532	1.052	0.800	0.726	0.726	0.540	0.818
	Entropy	0.688	0.545	1.014	0.818	0.745	0.745	0.508	0.842
Crowd-Text	*MV-BW*	0.665	0.602	2.133	0.749	0.722	0.722	0.648	0.784
	MV-DeepAE	0.682	0.611	1.372	0.792	0.738	0.738	0.615	0.800
	BCCwords	0.715	0.578	0.918	0.830	0.750	0.750	0.516	0.840
	CrowdDeepAE	**0.790**	**0.642**	**0.889**	**0.876**	**0.816**	**0.816**	**0.500**	**0.875**

Table 1: Comparison of crowdsourcing aggregation models on *CrowdFlower* (*CF*) and *SentimentPolarity* (*SP*) datasets, When 20% of crowd labels are available.

(a) Pos-docStatistic (b) Neg-docStatistic (c) Pos-CrowdDeepAE (d) Neg-CrowdDeepAE

Figure 3: Word clouds of the positive (Pos) and negative (Neg) sentiments in SP dataset using docStatistic and *CrowdDeepAE*.

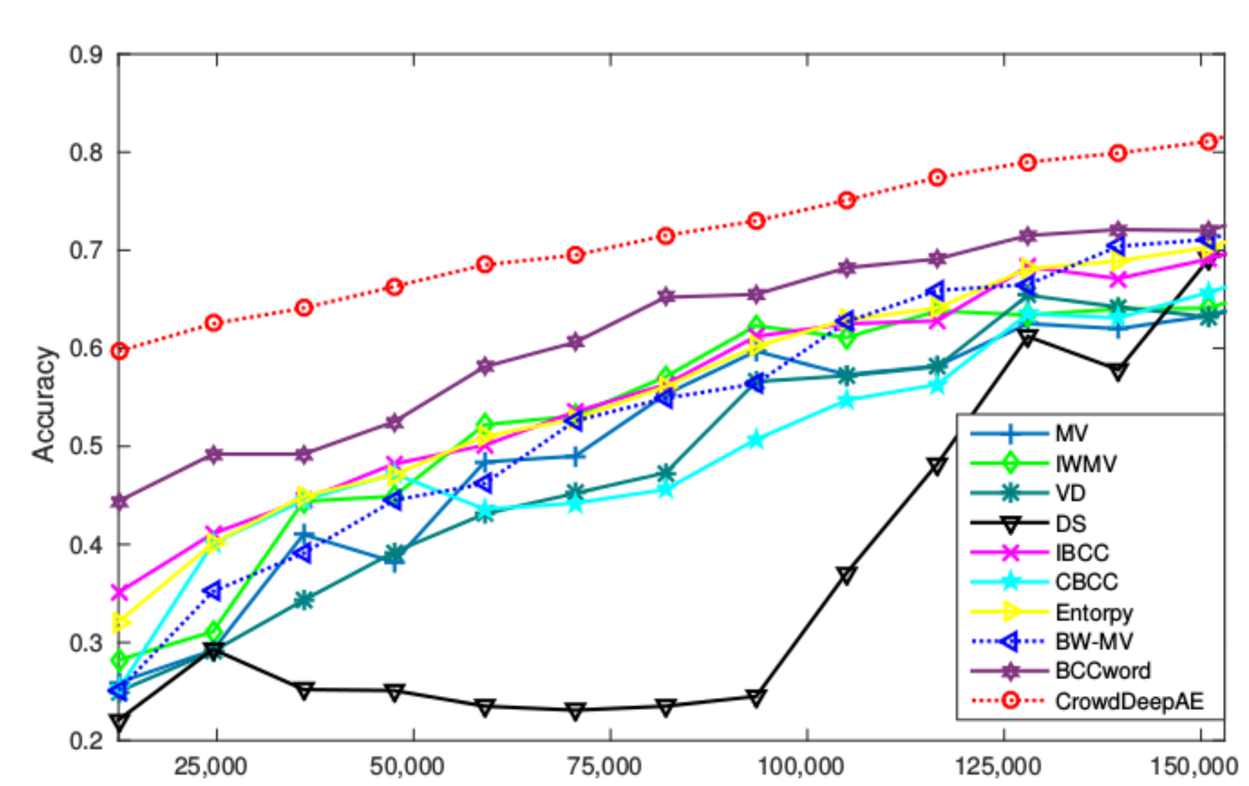
(a) Pos-docStatistic (b) Neg-docStatistic (c) Neut-docStatistic

(d) Pos-CrowdDeepAE (e) Neg-CrowdDeepAE (f) Neut-CrowdDeepAE

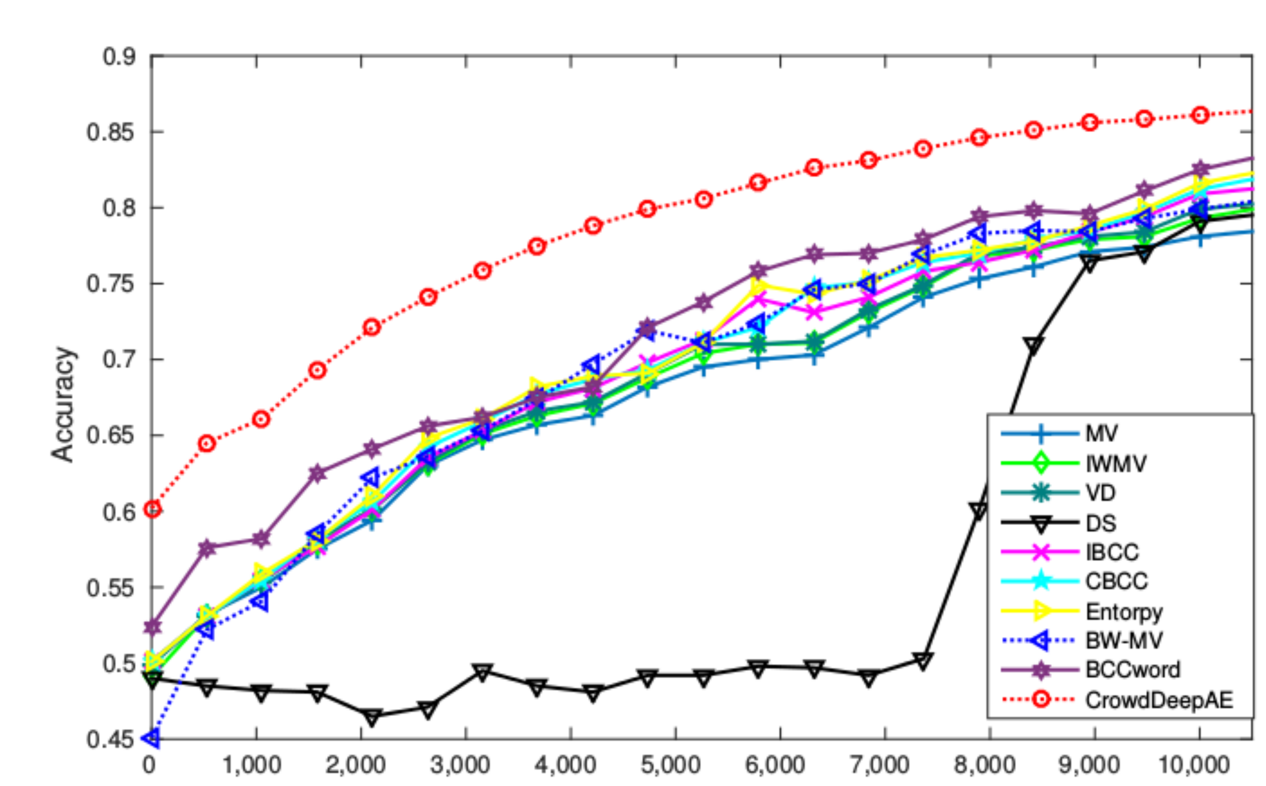
Figure 4: Word clouds of the positive (Pos), negative (Neg) and neutral (Neut) sentiments in SP dataset using docStatistic and CrowdDeepAE.

	Model	*CF* (all labels) Accuracy	Ave. recall	NLPD	AUC	*SP* (all labels) Accuracy	Ave. recall	NLPD	AUC
Crowd	*MV*	0.840	0.764	0.921	0.852	0.852	0.852	0.797	0.885
	IWMV	0.860	0.764	0.912	0.041	0.885	0.885	0.752	0.891
	VD	0.883	0.779	0.458	0.942	0.887	0.887	**0.338**	0.947
	DS	0.830	0.745	0.459	0.897	0.914	0.914	0.340	**0.957**
	IBCC	0.860	0.763	**0.437**	0.935	**0.915**	**0.915**	0.374	**0.957**
	CBCC	0.886	0.746	0.526	0.942	**0.915**	**0.915**	0.383	**0.957**
	Entropy	0.886	0.746	0.551	0.938	0.914	0.914	0.391	**0.957**
Crowd-Text	*MV-BW*	0.867	0.764	0.921	0.859	0.885	0.885	0.797	0.891
	MV-DeepAE	0.880	0.768	0.571	0.922	0.885	0.885	0.752	0.891
	BCCwords	0.890	0.807	0.591	0.877	**0.915**	**0.915**	0.389	**0.957**
	CrowdDeepAE	**0.912**	**0.825**	0.479	**0.948**	**0.915**	**0.915**	0.389	**0.957**

Table 2: Comparison of crowdsourcing aggregation models on *CrowdFlower* (*CF*) and *SentimentPolarity* (*SP*) datasets, When all crowd labels are available.



(a) *CrowdFlower* (*CF*)



(b) *SentimentPolarity* (*CF*)

Figure 5: Accuracy of crowdsourcing aggregation models on *CrowdFlower* (*CF*) and *SentimentPolarity* (SP) datasets, when increasing the number of crowd labels.