Appendix: Multi-label Categorization of Accounts of Sexism using a Neural Framework

Pulkit Parikh[1], **Harika Abburi**[1], **Pinkesh Badjatiya**[1], **Radhika Krishnan**[1],
Niyati Chhaya[1,2], **Manish Gupta**[1*], **Vasudeva Varma**[1]

[1]IIIT-Hyderabad, India
[2]Adobe Research, India
{pulkit.parikh,harika.a,pinkesh.badjatiya}@research.iiit.ac.in,
radhika.krishnan@iiit.ac.in, nchhaya@adobe.com, {manish.gupta,vv}@iiit.ac.in

1 Evaluation Metrics

The standard metrics used in single-label multi-class classification are inapplicable for our multi-label classification problem. We adopt established example (instance) based metrics, namely F1 (F_I) and accuracy (Acc_I), and label-based metrics, namely F1 macro (F_{macro}) and F1 micro (F_{micro}), defined as follows.

$$Acc_I = \frac{1}{n} \sum_{i=1}^{n} \frac{|\mathbf{y_i^+} \cap \hat{\mathbf{y}}_{\mathbf{i}}^+|}{|\mathbf{y_i^+} \cup \hat{\mathbf{y}}_{\mathbf{i}}^+|}; F_I = \frac{2P_I R_I}{P_I + R_I} \quad (1)$$

where

$$P_{inst} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\mathbf{y_i^+} \cap \hat{\mathbf{y}}_{\mathbf{i}}^+|}{|\hat{\mathbf{y}}_{\mathbf{i}}^+|}; R_{inst} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\mathbf{y_i^+} \cap \hat{\mathbf{y}}_{\mathbf{i}}^+|}{|\mathbf{y_i^+}|}$$

$$F_{macro} = \frac{1}{L} \sum_{j=1}^{L} F(TP_j, FP_j, FN_j) \quad (2)$$

$$F_{micro} = F\left(\sum_{j=1}^{L} TP_j, \sum_{j=1}^{L} FP_j, \sum_{j=1}^{L} FN_j\right) \quad (3)$$

where

$$TP_j = |\{x_i \mid l_j \in (\mathbf{y_i^+} \cap \hat{\mathbf{y}}_{\mathbf{i}}^+), 1 \le i \le n\}|, \quad (4)$$

$$FP_j = |\{x_i \mid l_j \in (\hat{\mathbf{y}}_{\mathbf{i}}^+ - \mathbf{y_i^+}), 1 \le i \le n\}|, \quad (5)$$

$$FN_j = |\{x_i \mid l_j \in (\mathbf{y_i^+} - \hat{\mathbf{y}}_{\mathbf{i}}^+), 1 \le i \le n\}|, \quad (6)$$

$$F(TP^*, FP^*, FN^*) = \frac{2TP^*}{2TP^* + FN^* + FP^*} \quad (7)$$