Here we give a more rigorous mathematical treatment of the learning methods presented in section 3 of the paper.

Method M_1: Direct Probability Estimation

Given a set $J_{k,i}$, we may define the parameter set Φ that has a single smoothing parameter

$$\Phi = \{\tau\} \tag{0.1}$$

Given the training set $\Xi(J_{k,i})$ the method $M_1(\Phi)$ defines the probability distribution of a query-document pair $dp \in DP$.

$$P_{dp}\left(c \mid M_1(\Phi), \Xi(J_{k,i})\right) = \frac{\sum_{j \in J_{k,i}} \delta\left(\zeta_j^{dp} - c\right) + \Theta_{k,i}(c) \cdot \tau}{11 + 11 \cdot \tau} \tag{0.2}$$

where

$$\delta(X) = \begin{cases} 1 \text{ , if } X = 0 \\ 0, \text{ otherwise} \end{cases} \tag{0.3}$$

and the quantity $\Theta_{k,i}(c)$ is the expected frequency of the class label $c \in C$ for the training set $\Xi(J_{k,i})$, and is given by

$$\Theta_{k,i}(c) = \frac{1}{|DP|} \sum_{dp \in DP} \sum_{j \in J_{k,i}} \delta\left(\zeta_j^{dp} - c\right). \tag{0.4}$$

Applying (0.2) to the maximization process in(5), one can induce the optimal parameter τ_k^* for a given test judge k. Given the optimal parameter τ_k^* and the training set $\Xi(J_k)$, the method M_1 provides the probability of the class label of a query-document pair $dp \in DP$ as follows:

$$P_{dp}\left(c \mid M_1(\Phi_k^*), \Xi(J_k)\right) = \frac{\sum_{j \in J_k} \delta\left(\zeta_j^k - c\right) + \Theta_k(c) \cdot \tau_k^*}{12 + 12 \cdot \tau_k^*} \tag{0.5}$$

where

$$\Theta_k(c) = \frac{1}{|DP|} \sum_{dp \in DP} \sum_{j \in J_k} \delta\left(\zeta_j^{dp} - c\right). \tag{0.6}$$

We can then use (0.5) for evaluation in formula (6).

Method M_2: Direct Probability Estimation With Weighting Parameters. It is not optimal to put each judge on an equal footing for his class label judgments of query- document pairs as the previous method M_1 does since the quality or reliability of judgments will differ among judges. Given a judge k, let us set

$$\Phi_k = \left\{ w_{j,c} \right\}_{j \in J_k}^{c \in C} \cup \left\{ \tau \right\}. \tag{0.7}$$

Incorporating the parameter $w_{j,c}$ for the class label c made by the judge j into the method M_1, the method M_2 defines the probability distribution over the class variable c

$$P_{dp}\left(c \mid M_2\left(\Phi_k\right), \Xi\left(J_{k,i}\right)\right) = \frac{\sum_{j \in J_{k,i}} \delta\left(\zeta_j^{dp} - c\right) \exp\left(w_{j,c}\right) + \Theta_{k,i}\left(c\right) \cdot \tau_k^*}{\sum_{c' \in C} \sum_{j \in J_{k,i}} \delta\left(\zeta_j^{dp} - c'\right) \exp\left(w_{j,c'}\right) + 11 \cdot \tau_k^*} \tag{0.8}$$

where τ_k^* is the smoothing factor obtained from the method M_1. Here we lose no generality by using τ_k^* as determined for M_1 and the same k. Just as for M_1 we may apply (0.8) in (5) to induce the optimal Φ_k^*. Then we may set

$$P_{dp}\left(c \mid M_2\left(\Phi_k^*\right), \Xi\left(J_k\right)\right) = \frac{\sum_{j \in J_k} \delta\left(\zeta_j^{dp} - c\right) \exp\left(w_{j,c}^*\right) + \Theta_k\left(c\right) \cdot \tau_k^*}{\sum_{c' \in C} \sum_{j \in J_k} \delta\left(\zeta_j^{dp} - c'\right) \exp\left(w_{j,c'}^*\right) + 12 \cdot \tau_k^*} \tag{0.9}$$

and use (0.9) for evaluation in formula (6).

Method M_3: Correlation Matrix with Weighting Parameters. Given a set $J_{k,i}$ and a query document pair $dp \in DP$, let us define the correlation matrix for a judge $j \in J_{k,i}$

$$\left(\boldsymbol{\lambda}_{k,i}^{j}\right)_{m,n}^{dp} = \sum_{l \in J_{k,i,j}} \delta\left(\zeta_j^{dp} - m\right) \delta\left(\zeta_l^{dp} - n\right) \text{ where } m, n \in C. \tag{0.10}$$

We may sum up (0.10) over all query-document pairs in the set DP

$$\left(\boldsymbol{\lambda}_{k,i}^{j}\right)_{m,n} = \sum_{dp \in DP} \left(\boldsymbol{\lambda}_{k,i}^{j}\right)_{m,n}^{dp}. \tag{0.11}$$

Then normalizing each row in the matrix (0.11), we get the following normalized correlation matrix

$$\left(\Lambda_{k,i}^{j}\right)_{m,n} = \frac{\left(\boldsymbol{\lambda}_{k,i}^{j}\right)_{m,n}}{\sum_{l\in C}\left(\boldsymbol{\lambda}_{k,i}^{j}\right)_{m,l}}. \tag{0.12}$$

We may view $\left(\Lambda_{k,i}^{j}\right)_{m,n}$ as the probability of the class label being n when the class label m is chosen by judge j. Because judges differ in their reliability, let us introduce the parameter $v_{j,c}$ for the purpose of weighting row $c \in C$ of the correlation matrix $\Lambda_{k,i}^{j}$ corresponding to judge $j \in J_{k,i}$. Then the method M_3 defines the probability distribution for a query- document pair $dp \in DP$ by

$$P_{dp}\left(c \mid M_3\left(\Phi_k\right), \Xi\left(J_{k,i}\right)\right) = \frac{\sum_{j\in J_{k,i}}\sum_{m\in C}\delta\left(m-\zeta_j^{dp}\right)\exp\left(v_{j,m}\right)\left(\Lambda_{k,i}^{j}\right)_{m,c} + \Theta_{k,i}\left(c\right)\tau}{\sum_{j\in J_{k,i}}\sum_{c'\in C}\sum_{m\in C}\delta\left(m-\zeta_j^{dp}\right)\exp\left(v_{j,m}\right)\left(\Lambda_{k,i}^{j}\right)_{m,c'} + 11\tau} \tag{0.13}$$

where

$$\Phi_k = \left\{v_{j,c}\right\}_{j\in J_k}^{c\in C} \cup \left\{\tau\right\}. \tag{0.14}$$

We then determine Φ_k^* in the standard way and apply the result to obtain

$$P_{dp}\left(c \mid M_3\left(\Phi_k^*\right), \Xi\left(J_k\right)\right) = \frac{\sum_{j\in J_k}\sum_{m\in C}\delta\left(\zeta_j^{dp}-m\right)\exp\left(v_{j,m}^*\right)\left(\Lambda_k^{j}\right)_{m,c} + \Theta_k\left(c\right)\tau_k^*}{\sum_{j\in J_k}\sum_{c'\in C}\sum_{m\in C}\delta\left(\zeta_j^{dp}-m\right)\exp\left(v_{j,m}^*\right)\left(\Lambda_k^{j}\right)_{m,c'} + 12\tau_k^*} \tag{0.15}$$

where the correlation matrix Λ_k^j can be obtained removing the test judge k only

$$\begin{aligned}
\left(\boldsymbol{\lambda}_k^j\right)_{m,n}^{dp} &= \sum_{l\in J_{k,j}} \delta\left(\zeta_j^{dp}-m\right)\cdot\delta\left(\zeta_l^{dp}-n\right) \\
\left(\boldsymbol{\lambda}_k^j\right)_{m,n} &= \sum_{dp\in DP}\left(\boldsymbol{\lambda}_k^j\right)_{m,n}^{dp} \\
\left(\Lambda_k^j\right)_{m,n} &= \frac{\left(\boldsymbol{\lambda}_k^j\right)_{m,n}}{\sum_{l\in C}\left(\boldsymbol{\lambda}_k^j\right)_{m,l}}.
\end{aligned} \tag{0.16}$$

Method M_{23}: Combining the methods M_2 and M_3.

We can combine the methods M_2 and M_3 defining the probability distribution over a class label c for a query-document pair dp by

$$P_{dp}\left(c \mid M_3\left(\Phi_k\right), \Xi\left(J_{k,i}\right)\right) = \frac{\sum_{j \in J_{k,i}} \left\{\delta\left(\zeta_j^{dp} - c\right)\exp\left(w_{j,c}\right) + \sum_{m \in C} \delta\left(m - \zeta_j^{dp}\right)\exp\left(v_{j,m}\right)\left(\Lambda_{k,i}^{j}\right)_{m,c}\right\} + \Theta_{k,i}\left(c\right)\tau}{\sum_{j \in J_{k,i}} \sum_{c' \in C} \left\{\delta\left(\zeta_j^{dp} - c'\right)\exp\left(w_{j,c'}\right) + \sum_{m \in C} \delta\left(m - \zeta_j^{dp}\right)\exp\left(v_{j,m}\right)\left(\Lambda_{k,i}^{j}\right)_{m,c'}\right\} + 11\tau} \quad (0.17)$$

where

$$\Phi_k = \left\{w_{j,c}\right\}_{j \in J_k}^{c \in C} \cup \left\{v_{j,c}\right\}_{j \in J_k}^{c \in C} \cup \left\{\tau\right\}. \quad (0.18)$$

We then determine Φ_k^* in the standard way and complete the definition of the method analogous to M_2 and M_3.

Method M_4: Intrinsic Judgments from a Weighted Average.

Yu et. al. devised the method whereby an intrinsic value (judgment) can be obtained from a suitably weighted average over judgments for any given item. Given the judge set $J_{k,i}$ and a query- document pair $dp \in DP$, one defines the weighted average of judgments of a query document pair $dp \in DP$ by

$$\mu_{k,i}^{dp} = \sum_{j \in J_{k,i}} r_j \zeta_j^{dp}. \quad (0.19)$$

Here the numbers r_j are a nonnegative normalized set and are designed to reflect the importance of each judges judgments. A judge's judging capability is reflected in the average quadratic error in her judging history on all query-document pairs in DP:

$$e_j^2 = \frac{1}{\left|DP\right|} \sum_{dp \in DP} \left(\zeta_j^{dp} - \mu_k^{dp}\right)^2 \text{ for any } j \in J_{k,i}. \quad (0.20)$$

Then the weights may be defined by

$$r_j = \frac{1/e_j}{\sum_{j' \in J_{k,i}} 1/e_{j'}}. \quad (0.21)$$

Starting with uniform weighting, the algorithm iterates eqs. (0.19), (0.20), and (0.21) to convergence to a solution. Once the solution has been obtained and we have the intrinsic class values $\bar{\mu}_{k,i}^{dp}$ of query-document pairs, we can define a probability distribution

$$P_{dp}\left(c \mid M_4\left(\Phi_k\right), \Xi\left(J_{k,i}\right)\right) = \exp\left(-\frac{\left(c-\bar{\mu}_{k,i}^{dp}\right)^2}{2\sigma_k^2}\right) \Bigg/ \sum_{c' \in C} \exp\left(-\frac{\left(c'-\bar{\mu}_{k,i}^{dp}\right)^2}{2\sigma_k^2}\right) \qquad (0.22)$$

where $\Phi_k = \{\sigma_k\}$. We then employ (0.22) to implement the optimization process (5) and induce the optimal $\Phi_k^* = \{\sigma_k^*\}$. Next we replace $J_{k,i}$ by J_k in the equations (0.19), (0.20), and (0.21) and solve to obtain the intrinsic values μ_k^{dp}. We can then define

$$P_{dp}\left(c \mid M_4\left(\Phi_k^*\right), \Xi\left(J_k\right)\right) = \exp\left(-\frac{\left(c-\bar{\mu}_k^{dp}\right)^2}{2\left(\sigma_k^*\right)^2}\right) \Bigg/ \sum_{c' \in C} \exp\left(-\frac{\left(c'-\bar{\mu}_k^{dp}\right)^2}{2\left(\sigma_k^*\right)^2}\right) \qquad (0.23)$$

and use (0.23) for evaluation as defined in (6).

Method M_5: Maximum Entropy Classifier.

For details of the Maximum Entropy classifier we refer the reader to (Berger, Pietra, and Pietra, 1996)[30]. Here data points to be classified correspond to the query-document pairs $dp \in DP$. In order to apply a maximum entropy classifier we need to define a class label and features for each instance. Given a judge set $J_{k,i}$ and a query- document pair $dp \in DP$, let us define an instance corresponding to a judge $j \in J_{k,i}$ as

$$\boldsymbol{\omega}_{k,i}^{dp}(j) = \left(\left\{\left(l, \zeta_l^{dp}\right)\right\}_{l \in J_{k,i,j}}, \zeta_j^{dp}\right) \qquad (0.24)$$

where the first coordinate, $\left\{\left(l, \zeta_l^{dp}\right)\right\}_{l \in J_{k,i,j}}$, is the set of features and the second coordinate, ζ_j^{dp}, is the class label for the instance. The complete set of instances may be represented as

$$\boldsymbol{\Omega}_{k,i} = \left\{\boldsymbol{\omega}_{k,i}^{dp}(j)\right\}_{j \in J_{k,i}}^{dp \in DP}. \qquad (0.25)$$

Now the maximum entropy classifier requires a regularization parameter λ, so for us $\Phi_k = \{\lambda_k\}$. In order to define the feature functions required by the maximum entropy method we let $\omega \in \Omega_{k,i}$ and allow ω_1 to represent the first coordinate or set of features for the instance and ω_2 the second coordinate or label of the instance. Then for each pair (j,m) where $j \in J_{k,i}$ and $m \in C$ and for each $c \in C$ there is a feature function $f_{(j,m)}^{c}$ defined by

$$f_{(j,m)}^{c}(\omega_1,\omega_2) = \begin{cases} 1 \text{ if } (j,m) \in \omega_1 \text{ and } c = \omega_2 \\ 0 \text{ otherwise} \end{cases}. \qquad (0.26)$$

Training the classifier on $\Omega_{k,i}$ will produce a set of weights

$$\left\{ \alpha_{c}^{(j,m)} \right\}. \qquad (0.27)$$

We can use these weights to estimate label probabilities for any object whose features are appropriate. We apply them to the object dp with the set of features $\eta(dp,k,i) = \left\{ \left(j, \zeta_{j}^{dp} \right) \right\}_{j \in J_{k,i}}$ and obtain the probability estimates

$$P_{dp}\left(c \mid M_5(\Phi_k), \Xi(J_{k,i}) \right) = \frac{1}{Z} \exp\left(\sum_{(j,m)\in\eta(dp,k,i), c' \in C} \alpha_{c'}^{(j,m)} \cdot f_{(j,m)}^{c'}\left(\eta(dp,k,i), c \right) \right) \qquad (0.28)$$

where Z is the standard normalizing factor.

Applying (0.28) to the maximization process (5), we can induce the optimal parameter $\Phi^* = \left\{ \lambda_k^* \right\}$ for given test judge k. Given the optimal parameter λ_k^* , were rewrite (0.24) as

$$\boldsymbol{\omega}_k^{dp}(j) = \left(\left\{ \left(l, \zeta_l^{dp} \right) \right\}_{l \in J_{k,j}}, \zeta_j^{dp} \right) \qquad (0.29)$$

and (0.25)

$$\boldsymbol{\Omega}_k = \left\{ \boldsymbol{\omega}_k^{dp}(j) \right\}_{j \in J_k}^{dp \in DP} \qquad (0.30)$$

and train the maximum entropy classifier on Ω_k using λ_k^*. We then redefine the feature set for the object dp by $\eta(dp,k) = \left\{ \left(j, \zeta_j^{dp} \right) \right\}_{j \in J_k}$ and in analogy with (0.28) set

$$P_{dp}\left(c \mid M_5(\Phi_k^*), \Xi(J_k) \right) = \frac{1}{Z} \exp\left(\sum_{(j,m)\in\eta(dp,k), c' \in C} \alpha_{c'}^{(j,m)} \cdot f_{(j,m)}^{c'}\left(\eta(dp,k), c \right) \right) \qquad (0.31)$$

where $\alpha_c^{(j,m)}$ are the newly learned weights based on (0.30). We then use (0.31) for evaluation as defined in (6).