IRT-based Aggregation Model of Crowdsourced Pairwise Comparisons for Evaluating Machine Translations: Appendix

Naoki Otani[1] **Toshiaki Nakazawa**[2] **Daisuke Kawahara**[1] **Sadao Kurohashi**[1]
[1]Graduate School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, Japan
[2]Japan Science and Technology Agency, Kawaguchi-shi, Saitama, Japan
otani.naoki.65v@st.kyoto-u.ac.jp nakazawa@pa.jst.jp {dk,kuro}@i.kyoto-u.ac.jp

Abstract

This document complements the main paper "IRT-based Aggregation Model of Crowdsourced Pairwise Comparisons for Evaluating Machine Translations". We follow the same notations presented in the main paper.

A Generative Judgment Model

The generative probability of judgment $u_{i,j,k}$ is defined as he difference in the BCC, that is,

$$\begin{aligned}\mathrm{P}_{jkc}(\theta_i) &= \mathrm{P}(u_{i,j,k}=c|\theta_i, b_j, a_k)\\ &= \mathrm{P}^*_{jkc-1}(\theta_i) - \mathrm{P}^*_{jkc}(\theta_i).\end{aligned}$$

Thus, they can be written as,

$$\begin{aligned}\mathrm{P}(u_{i,j,k}=1|\theta_i, b_j, a_k) &= 1 - \frac{1}{1+\exp\left(-a_k(\theta_i - b_{j1})\right)}\\ \mathrm{P}(u_{i,j,k}=2|\theta_i, b_j, a_k) &= \frac{1}{1+\exp\left(-a_k(\theta_i - b_{j1})\right)}\\ &\quad - \frac{1}{1+\exp\left(-a_k(\theta_i - b_{j2})\right)}\\ \mathrm{P}(u_{i,j,k}=3|\theta_i, b_j, a_k) &= \frac{1}{1+\exp\left(-a_k(\theta_i - b_{j2})\right)}.\end{aligned}$$

B Parameter Estimation

We find the values of the parameters to maximize the log likelihood based on obtained judgments U.

$$\mathcal{L}(\theta, \xi) = \log\mathrm{P}(U, \theta, \xi)$$

B.1 Marginal Likelihood Maximization of Judge Sensitivity and Matchup Difficulty

We first estimate the parameters ξ to maximize the marginal log likelihood w.r.t. the system performance θ.

$$\begin{aligned}m\mathcal{L}(\xi) &= \log\mathrm{P}(U, \xi)\\ &= \sum_{i\in\mathcal{I}} \log \int_{-\infty}^{\infty} \mathrm{P}(\theta)\mathrm{P}(U_i|\theta,\xi)d\theta + \log\mathrm{P}(\xi).\end{aligned}$$

The equation above can be approximated using Gauss-Hermite quadrature, i.e.,

$$\begin{aligned}m\mathcal{L}(\xi) &\approx \sum_{i\in\mathcal{I}} \log \sum_{t=1}^{T} \frac{1}{\sqrt{\pi}} w_t \mathrm{P}(U_i|\tau x_t, \xi) + \log\mathrm{P}(\xi)\\ w_t &= \frac{2^{T-1}T!\sqrt{\pi}}{T^2\left(H(x_t)\right)^2}\\ H(x_t) &= \left(2x_t - \frac{d}{dx_t}\right)^{T-1} \cdot 1,\end{aligned}$$

where x_t and w_t were precomputed as shown in Table 1.

We solve the optimization problem using the gradient descent methods to maximize the approximated marginal likelihood.

$$\begin{aligned}\frac{\partial}{\partial a_k} m\mathcal{L} &= \sum_{i\in\mathcal{I}} \left(\tilde{\mathrm{P}}(U_i|\xi)\right)^{-1} \sum_{t=1}^{T} \frac{1}{\sqrt{\pi}} w_t \frac{\partial}{\partial a_k} \mathrm{P}(U_i|\tau x_t, \xi)\\ &\quad + \frac{\partial}{\partial a_k}\log\mathrm{P}(a_k)\\ &= \sum_{i\in\mathcal{I}} \sum_{t=1}^{T} W_{it} \frac{\partial}{\partial a_k}\log\mathrm{P}(U_i|\tau x_t, \xi) + \frac{\partial}{\partial a_k}\log\mathrm{P}(a_k),\end{aligned}$$

t	x_t	w_t
1	-5.5503518730	3.7203650700E-014
2	-4.7739923430	8.8186112000E-011
3	-4.1219955470	2.5712301800E-008
4	-3.5319728770	2.1718848980E-006
5	-2.9799912080	0.0000747840
6	-2.4535521250	0.0012549820
7	-1.9449629490	0.0114140658
8	-1.4489342510	0.0601796467
9	-0.9614996344	0.1921203241
10	-0.4794507071	0.3816690740
11	0.0000000000	0.4790237031
12	0.4794507071	0.3816690740
13	0.9614996344	0.1921203241
14	1.4489342510	0.0601796467
15	1.9449629490	0.0114140658
16	2.4535521250	0.0012549820
17	2.9799912080	0.0000747840
18	3.5319728770	2.1718848980E-006
19	4.1219955470	2.5712301800E-008
20	4.7739923430	8.8186112000E-011
21	5.5503518730	3.7203650700E-014

Table 1: Values of x_t and t_t used in the Gauss-Hermite quadrature.

where

$$\tilde{\mathrm{P}}(U_i|\xi) = \sum_{t=1}^{T} \frac{1}{\sqrt{\pi}} w_t \mathrm{P}(U_i|\tau x_t, \xi).$$

$$\begin{aligned} W_{it} &= \frac{w_t \mathrm{P}(U_i, \xi|\tau x_t)/\sqrt{\pi}}{\tilde{\mathrm{P}}(U_i, \xi)} \\ &= \frac{w_t \mathrm{P}(U_i|\tau x_t, \xi)}{\Sigma_{t'=1}^{T} w_{t'} \mathrm{P}(U_i|\tau x_{t'}, \xi)}. \end{aligned}$$

In the same way,

$$\begin{aligned} \frac{\partial}{\partial b_{jc}} m\mathcal{L} = \sum_{i \in \mathcal{I}} \sum_{t=1}^{T} W_{it} \frac{\partial}{\partial b_{jc}} \mathrm{logP}(U_i|\tau x_t, \xi) \\ + \frac{\partial}{\partial b_{jc}} \mathrm{P}(b_{jc}). \qquad (c = 1, 2) \end{aligned}$$

The inequality constraints on the parameters are handled by adding log barrier functions to the objective function.

B.2 Maximum A Posteriori (MAP) Estimation of System Performance

Given the estimates of ξ, we estimate the system performance θ by using MAP estimation. We maximize the objective function,

$$\begin{aligned} \mathcal{L}(\theta) &= \mathrm{logP}(U, \theta; \xi) \\ &= \sum_{i \in \mathcal{I}} \mathrm{logP}(\theta) + \sum_{i \in \mathcal{I}} \mathrm{logP}(U|\theta; \xi). \end{aligned}$$

The estimates of θ are obtained using the gradient descent method. The gradient for parameter θ_i ($i \in \mathcal{I}$) is

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta_i} &= -\frac{2\theta_i}{\tau^2} + \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} \sum_{c \in \mathcal{C}} \frac{u_{i,j,k,c}}{\mathrm{P}_{jkc}(\theta_i)} \frac{\partial}{\partial \theta_i} \mathrm{P}_{jkc}(\theta_i) \\ &= -\frac{2\theta_i}{\tau^2} + \sum_{j,k,c} \frac{u_{i,j,k,c}}{\mathrm{P}_{jkc}(\theta_i)} a_k (\mathrm{P}^*_{jkc} \mathrm{Q}^*_{jkc} - \mathrm{P}^*_{jkc+1} \mathrm{Q}^*_{jkc+1}), \end{aligned}$$

where $Q^* = 1 - P^*$.