GLUE: derivation of objective

Soel Micheletti

June 11, 2022

The goal is finding $Pr[u, V|x_k, \mathcal{G}] = \frac{Pr[u,V]Pr[x_k,\mathcal{G}|u,V]}{\int Pr[x_k,\mathcal{G}|u,V]Pr[u,v]dudv}$

Unfortunately, because of the integral in the denominator, the formula above is intractable. We hence introduce a posterior $q_{\phi_k\phi_G}[u, V|x_k, \mathcal{G}] = q_{\phi_k}[u|x_k]q_{\phi_G}[V|\mathcal{G}]$. The equality follows from the fact that the encoders for the graph and for the omics layers are independent from each other.

We want our posterior to be as close as possible to the original distribution that we want to approximate. This justifies using the following objective:

$$\arg\min_{\phi_k,\phi_G} KL[q_{\phi_k\phi_G}[u, V|x_k, \mathcal{G}] \quad || \quad Pr[u, V|x_k, \mathcal{G}]]$$

By rearranging the objective we get:

$$\begin{aligned} KL[q_{\phi_k\phi_G}[u, V|x_k, \mathcal{G}] \quad || \quad Pr[u, V|x_k, \mathcal{G}]] &= E_{u,v\sim q_{\phi_k\phi_G}[u,V|x_k,\mathcal{G}]}\left[\log\left(\frac{q_{\phi_k\phi_G}[u, V|x_k, \mathcal{G}]}{Pr[u, V|x_k, \mathcal{G}]}\right)\right] \\ &= E_{u,v\sim q_{\phi_k\phi_G}[u,V|x_k,\mathcal{G}]}\left[\log\left(q_{\phi_k\phi_G}[u, V|x_k, \mathcal{G}]\right) - \log\left(Pr[u, V|x_k, \mathcal{G}]\right)\right] \end{aligned}$$

By using the relation $Pr[u, V|x_k, \mathcal{G}] = \frac{Pr[u,V,x_k,\mathcal{G}]}{Pr[x_k,\mathcal{G}]}$, the above formula becomes:

$$E_{u,v\sim q_{\phi_k\phi_G}[u,V|x_k,\mathcal{G}]}\left[\log\left(q_{\phi_k\phi_G}[u, V|x_k, \mathcal{G}]\right)\right] - E_{u,v\sim q_{\phi_k\phi_G}[u,V|x_k,\mathcal{G}]}\left[\log\left(Pr[u, V, x_k, \mathcal{G})\right] + \log\left(Pr[x_k, \mathcal{G}]\right)\right.$$

Where we can ignore the last term because it is independent from the parameters. By flipping signs, we now have the following objective (equivalent to the original one).

$$\begin{aligned} &\arg\max_{\phi_k,\phi_G} E_{u,v\sim q_{\phi_k\phi_G}[u,V|x_k,\mathcal{G}]}\left[\log\left(Pr[u, V, x_k, \mathcal{G})\right] - E_{u,v\sim q_{\phi_k\phi_G}[u,V|x_k,\mathcal{G}]}\left[\log\left(q_{\phi_k\phi_G}[u, V|x_k, \mathcal{G}]\right)\right]\right. \\ =& \arg\max_{\phi_k,\phi_G} E_{u,v\sim q_{\phi_k\phi_G}[u,V|x_k,\mathcal{G}]}\left[\log\left(Pr[x_k, \mathcal{G}|u, V]Pr[u, V]\right)\right] - E_{u,v\sim q_{\phi_k\phi_G}[u,V|x_k,\mathcal{G}]}\left[\log\left(q_{\phi_k\phi_G}[u, V|x_k, \mathcal{G}]\right)\right] \\ =& \arg\max_{\phi_k,\phi_G} E_{u,v\sim q_{\phi_k\phi_G}[u,V|x_k,\mathcal{G}]}\left[\log\left(Pr[x_k, \mathcal{G}|u, V]\right)\right] - E_{u,v\sim q_{\phi_k\phi_G}[u,V|x_k,\mathcal{G}]}\left[\log\left(\frac{q_{\phi_k\phi_G}[u, V|x_k, \mathcal{G}]}{Pr[u, V]}\right)\right] \\ =& \arg\max_{\phi_k,\phi_G} E_{u,v\sim q_{\phi_k\phi_G}[u,V|x_k,\mathcal{G}]}\left[\log\left(Pr[x_k, \mathcal{G}|u, V]\right)\right] - KL\left[q_{\phi_k\phi_G}[u, V|x_k, \mathcal{G}]||Pr[u, V]\right] \end{aligned}$$

The term above is called Evidence Lower Bound (ELBO). Note that it can be simplified by using some (conditional) independencies. We will use that:

- The encoders are independent
- The graph decoder uses only the graph embedding

- The priors of the embeddings are independent

So the term above can be simplified to:

$$\begin{aligned}\arg\max_{\phi_k,\phi_G} E_{u\sim q_{\phi_k}[u|x_k],v\sim q_{\phi_G}[V|\mathcal{G}]}\left[\log\left(Pr[x_k|u,V]Pr[\mathcal{G}|V]\right)\right] - KL\left[q_{\phi_k}[u|x_k]q_{\phi_G}[V|\mathcal{G}]||Pr[u]Pr[V]\right] \\ = \arg\max_{\phi_k,\phi_G} E_{u\sim q_{\phi_k}[u|x_k],v\sim q_{\phi_G}[V|\mathcal{G}]}\left[\log\left(Pr[x_k|u,V]\right)\right] + E_{v\sim q_{\phi_G}[V|\mathcal{G}]}\left[\log\left(Pr[\mathcal{G}|V]\right)\right] \\ -KL\left[q_{\phi_k}[u|x_k]||Pr[u]\right] - KL\left[q_{\phi_G}[V|\mathcal{G}]||Pr[V]\right] \\ =: ELBO(x_k)\end{aligned}$$

We need to add two things:

- The decoders to parametrize $\log(Pr[x_k|u,V])$ and $\log(Pr[\mathcal{G}|V])$. For the data decoder, we use a linear decoder and some distribution (e.g. the negative binomial), while for the graph decoder a simple maximum likelihood with probability of having an edge proportional to the scalar product of the embeddings (so that we encourage linked features to have similar embeddings). So we will write $\log(Pr[x_k|u,V,\Theta_k])$ to indicate the parameters of the omics decoder. The graph decoder does not need to estimate any parameter.
- The expectation through all omics layers $1...K$.

Hence, the objective now becomes:

$$\arg\max_{\phi,\phi_G,\Theta}\sum_{k=1}^{K} E_{x_k\sim p_{\text{data}}(x_k)} ELBO(x_k)$$

which can be further rearranged into the following form

$$K\cdot\mathcal{L}_G(\phi_G) + \sum_{k=1}^{K}\mathcal{L}_{\mathcal{X}_k}(\Theta_k,\phi_k,\phi_G)$$

where we have

- $\mathcal{L}_G(\phi_G) = E_{V\sim q_{\phi_G}[V|\mathcal{G}]}\left[\log\left(Pr[\mathcal{G}|V]\right)\right] - KL\left[q_{\phi_G}[V|\mathcal{G}]||Pr[V]\right]$
- $\mathcal{L}_{\mathcal{X}_k}(\Theta_k,\phi_k,\phi_G) = E_{x_k\sim p_{\text{data}}(x_k)}\left[E_{u\sim q_{\phi_k}[u|x_k],v\sim q_{\phi_G}[V|\mathcal{G}]}\left[\log\left(Pr[x_k|u,V,\Theta_k]\right)\right] - KL\left[q_{\phi_k}[u|x_k]||Pr[u]\right]\right]$

The last missing step is the adversarial learning of the feature embeddings to ensure proper alignment. We introduce a decoder D that assigns the probability that a given embedding u is generated from the k-th omic layer. More precisely, we minimize

$$\mathcal{L}_D(\phi,\psi) = -\frac{1}{K}\sum_{k=1}^{K} E_{x_k\sim p_{\text{data}}(x_k)}\left[E_{u\sim q_{\phi_k}(u|x_k)}\left[logD_k(u,\psi)\right]\right]$$

Hence, the overall objective of GLUE is given by:

$$\begin{aligned}&\arg\min_{\psi}\lambda_D\mathcal{L}_D(\phi,\psi) \\ \arg\max_{\Theta,\phi}\lambda_D\mathcal{L}_D(\phi,\psi) + \lambda_G K\mathcal{L}_G(\phi_G) &+ \sum_{k=1}^{K}\mathcal{L}_{\mathcal{X}_k}(\Theta_k,\phi_k,\phi_G)\end{aligned}$$

where λ_D and λ_G control the contributions of alignment and graph-based feature embedding.