Supplement for the paper: Generative Modeling for Maximizing Precision and Recall in Information Visualization

Jaakko Peltonen[1] **Samuel Kaski**[1,2]
[1]Aalto University, Department of Information and Computer Science,
Helsinki Institute for Information Technology HIIT, P.O. Box 15400, FI-00076 Aalto, Finland
[2]University of Helsinki, Department of Computer Science,
Helsinki Institute for Information Technology HIIT, P.O. Box 68, 00014 University of Helsinki, Finland

Abstract

In this supplement we derive the precision-recall analysis discussed in the main paper for the mixture model likelihood (equations (6)-(7) in Section 2.1 of the main paper).

Consider a query point i, and let the set of its actually relevant neighbors be P_i. Assume that the user (or the retrieval model) retrieves a set of points R_i as neighbors based on the visualization. We may assume that $|P_i| \geq 1$ and $|R_i| \geq 1$. Assume that the user (or the retrieval model) places uniform high probabilities r_{ij} across the set of points R_i and very low probabilities for other points, and assume similarly that the probabilities p_{ij} have uniform high values across the set of points P_i and low values for other points, so that

$$p_{ij} = \begin{cases} a_i \equiv \frac{1-\epsilon}{|P_i|} \,, & \text{if } j \in P_i \\ b_i \equiv \frac{\epsilon}{N-|P_i|-1} \,, & \text{otherwise} \end{cases} \quad (1)$$

$$r_{ij} = \begin{cases} c_i \equiv \frac{1-\epsilon}{|R_i|} \,, & \text{if } j \in R_i \\ d_i \equiv \frac{\epsilon}{N-|R_i|-1} \,, & \text{otherwise} \end{cases} \quad (2)$$

where ϵ is a very small positive number and $N-1$ is the total number of points other than i.

The log-likelihood of our mixture model for a single query point i can then be written as

$$\begin{aligned} L(i) &= \sum_{j\neq i} p_{ij} \log q_{ij} = \sum_{j\neq i} p_{ij} \log\left(\frac{r_{ij}+\gamma p_{ij}}{1+\gamma}\right) \\ &= -\log(1+\gamma) + \sum_{j\neq i} p_{ij}\log(r_{ij}+\gamma p_{ij}) \\ &= -\log(1+\gamma) + \sum_{j\neq i,\; j\in P_i\cap R_i} a_i \log(c_i+\gamma a_i) \\ &\quad + \sum_{j\neq i,\; j\in P_i\cap R_i^c} a_i \log(d_i+\gamma a_i) \\ &\quad + \sum_{j\neq i,\; j\in P_i^c\cap R_i} b_i \log(c_i+\gamma b_i) \\ &\quad + \sum_{j\neq i,\; j\in P_i^c\cap R_i^c} b_i \log(d_i+\gamma b_i) \end{aligned} \quad (3)$$

where R_i^c and P_i^c denote complements of R_i and P_i.

Denoting the number of true positives by $N_{TP,i} = |P_i \cap R_i|$, the number of false positives by $N_{FP,i} = |R_i \cap P_i^c|$, the number of misses by $N_{MISS,i} = |P_i \cap R_i^c|$ and the number of true negatives by $N_{TN,i} = |P_i^c \cap R_i^c|$, the log-likelihood becomes[1]

Appearing in Proceedings of the 14^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2011, Fort Lauderdale, FL, USA. Volume 15 of JMLR: W&CP 15.

[1]We assume $|P_i| < N-1$ and $|R_i| < N-1$. The special cases where $|P_i| = N-1$ or $|R_i| = N-1$ are briefly treated at the end of the proof, yielding the same results.

$$
\begin{aligned}
L(i) = & -\log(1+\gamma) + N_{TP,i}\cdot a_i \log(c_i+\gamma a_i) \\
& + N_{MISS,i}\cdot a_i \log(d_i+\gamma a_i) \\
& + N_{FP,i}\cdot b_i \log(c_i+\gamma b_i) + N_{TN,i}\cdot b_i \log(d_i+\gamma b_i) \\
= & -\log(1+\gamma) + N_{TP,i}\frac{1-\epsilon}{|P_i|}\log\left(\frac{1-\epsilon}{|R_i|}+\gamma\frac{1-\epsilon}{|P_i|}\right) \\
& + N_{MISS,i}\frac{1-\epsilon}{|P_i|}\log\left(\frac{\epsilon}{N-|R_i|-1}+\gamma\frac{1-\epsilon}{|P_i|}\right) \\
& + \frac{N_{FP,i}\cdot\epsilon}{N-|P_i|-1}\log\left(\frac{1-\epsilon}{|R_i|}+\gamma\frac{\epsilon}{N-|P_i|-1}\right) \\
& + \frac{N_{TN,i}\cdot\epsilon}{N-|P_i|-1}\log\left(\frac{\epsilon}{N-|R_i|-1}+\gamma\frac{\epsilon}{N-|P_i|-1}\right) \ .
\end{aligned} \tag{4}
$$

Rearranging terms, and noting that $N_{TP,i}+N_{MISS,i}=|P_i|$, equation (4) becomes

$$
\begin{aligned}
L(i) = & \frac{N_{TP,i}}{|P_i|}(1-\epsilon)\log\left((1-\epsilon)\left(\frac{1}{|R_i|}+\frac{\gamma}{|P_i|}\right)\right) \\
& + \frac{|P_i|-N_{TP,i}}{|P_i|}(1-\epsilon)\log\left(\frac{\gamma(1-\epsilon)}{|P_i|}+\frac{\epsilon}{N-|R_i|-1}\right) \\
& + \frac{N_{FP,i}}{N-|P_i|-1}\epsilon\log\left(\frac{1}{|R_i|}+\epsilon\left(\frac{\gamma}{N-|P_i|-1}-\frac{1}{|R_i|}\right)\right) \\
& + \frac{N_{TN,i}}{N-|P_i|-1}\Bigg[\epsilon\log\epsilon \\
& + \epsilon\log\left(\frac{1}{N-|R_i|-1}+\frac{\gamma}{N-|P_i|-1}\right)\Bigg] \\
& - \log(1+\gamma) \ .
\end{aligned} \tag{5}
$$

If ϵ is close enough to zero, only the first two lines on the right-hand side (arising from true positives and misses) contribute strongly to the cost function. The term $-\log(1+\gamma)$ is constant with respect to the visualization neighborhood R_i and all other terms are negligible in magnitude. We can thus write

$$
\begin{aligned}
L(i) \approx & -\log(1+\gamma) \\
& + \frac{N_{TP,i}}{|P_i|}(1-\epsilon)\log\left((1-\epsilon)\left(\frac{1}{|R_i|}+\frac{\gamma}{|P_i|}\right)\right) \\
& + \frac{|P_i|-N_{TP,i}}{|P_i|}(1-\epsilon)\log\left(\frac{\gamma(1-\epsilon)}{|P_i|}+\frac{\epsilon}{N-|R_i|-1}\right) \\
= & -\log(1+\gamma)+\frac{N_{TP,i}}{|P_i|}(1-\epsilon)\log\left(\frac{(1-\epsilon)\left(\frac{1}{|R_i|}+\frac{\gamma}{|P_i|}\right)}{\frac{\gamma(1-\epsilon)}{|P_i|}+\frac{\epsilon}{N-|R_i|-1}}\right) \\
& + (1-\epsilon)\log\left(\frac{\gamma(1-\epsilon)}{|P_i|}+\frac{\epsilon}{N-|R_i|-1}\right) \\
= & -\log(1+\gamma)+\frac{N_{TP,i}}{|P_i|}(1-\epsilon)\log\left(\frac{\frac{|P_i|}{|R_i|}+\gamma}{\gamma+\frac{\epsilon}{1-\epsilon}\cdot\frac{|P_i|}{N-|R_i|-1}}\right) \\
& + (1-\epsilon)\log\left(\frac{\gamma(1-\epsilon)}{|P_i|}+\frac{\epsilon}{N-|R_i|-1}\right) \ .
\end{aligned} \tag{6}
$$

Note that $N_{TP,i}/|P_i| = \mathit{recall}$. We also note that when $N_{TP,i} > 0$, we have $|P_i|/|R_i| = (N_{TP,i}/|R_i|)(|P_i|/N_{TP,i}) = \mathit{precision}/\mathit{recall}$, and $|P_i|/(N-|R_i|-1) = ((N-1)/|P_i| - |R_i|/|P_i|)^{-1} = (a-\mathit{recall}/\mathit{precision})^{-1}$ where $a=(N-1)/|P_i|$. The right-hand side of equation (6) can then be written as

$$
\begin{aligned}
& -\log(1+\gamma) \\
& + \mathit{recall}\cdot(1-\epsilon)\log\left(\frac{\frac{\mathit{precision}}{\mathit{recall}}+\gamma}{\gamma+\frac{\epsilon}{1-\epsilon}\cdot\left(a-\frac{\mathit{recall}}{\mathit{precision}}\right)^{-1}}\right) \\
& + (1-\epsilon)\log\left(\frac{\gamma(1-\epsilon)}{|P_i|}+\frac{\epsilon}{N-|R_i|-1}\right) \\
& = -\log(1+\gamma) \\
& + \mathit{recall}\cdot(1-\epsilon)\log\left(\frac{\left(\frac{\mathit{precision}}{\mathit{recall}}+\gamma\right)\left(a-\frac{\mathit{recall}}{\mathit{precision}}\right)}{\gamma\left(a-\frac{\mathit{recall}}{\mathit{precision}}\right)+\frac{\epsilon}{1-\epsilon}}\right) \\
& + (1-\epsilon)\log\left(\frac{\gamma(1-\epsilon)}{|P_i|}+\frac{\epsilon}{N-|R_i|-1}\right) \\
& \approx -\log(1+\gamma) \\
& + \mathit{recall}\cdot\log\left(\frac{\left(\frac{\mathit{precision}}{\mathit{recall}}+\gamma\right)\left(a-\frac{\mathit{recall}}{\mathit{precision}}\right)}{\gamma\left(a-\frac{\mathit{recall}}{\mathit{precision}}\right)+\frac{\epsilon}{1-\epsilon}}\right) \\
& + (1-\epsilon)\log\left(\frac{\gamma(1-\epsilon)}{|P_i|}+\frac{\epsilon}{N-|R_i|-1}\right)
\end{aligned} \tag{7}
$$

which corresponds to equation (6) in the main paper since $-\log(1+\gamma)$ is a constant. If ϵ is much smaller than γ, equation (7) simplifies to

$$
\begin{aligned}
L(i) &\approx -\log(1+\gamma) \\
&+ \mathit{recall} \cdot \log\left(\frac{\left(\frac{\mathit{precision}}{\mathit{recall}}+\gamma\right)\left(a-\frac{\mathit{recall}}{\mathit{precision}}\right)}{\gamma\left(a-\frac{\mathit{recall}}{\mathit{precision}}\right)}\right) \\
&+ (1-\epsilon)\log\left(\frac{\gamma(1-\epsilon)}{|P_i|}\right) \\
&= \mathit{const.} + \mathit{recall}\cdot\log\left(1+\frac{1}{\gamma}\cdot\frac{\mathit{precision}}{\mathit{recall}}\right) \qquad (8)
\end{aligned}
$$

which is the result in the main paper (equation (7) in the main paper).

On the other hand, if γ is much smaller than ϵ, equation (7) instead simplifies to

$$
\begin{aligned}
L(i) &\approx -\log(1+\gamma) \\
&+ \mathit{recall} \cdot \log\left(\frac{\left(\frac{\mathit{precision}}{\mathit{recall}}+\gamma\right)\left(a-\frac{\mathit{recall}}{\mathit{precision}}\right)}{\frac{\epsilon}{1-\epsilon}}\right) \\
&+ (1-\epsilon)\log\left(\frac{\epsilon}{N-|R_i|-1}\right) \\
&= \mathit{const.} + \mathit{recall}\cdot\log\left(\frac{\mathit{precision}}{\mathit{recall}}+\gamma\right) \\
&+ \mathit{recall}\cdot\log\left(a-\frac{\mathit{recall}}{\mathit{precision}}\right) \\
&+ \mathit{recall}\cdot\log\left(\frac{1-\epsilon}{\epsilon}\right) \\
&+ (1-\epsilon)\log\epsilon - (1-\epsilon)\log\left(N-|R_i|-1\right) \\
&\approx \mathit{const.} + \mathit{recall}\cdot\log\left(\frac{1-\epsilon}{\epsilon}\right) \\
&= \mathit{const.} + \mathit{recall}\cdot\mathit{const.} \qquad (9)
\end{aligned}
$$

where the approximation on the second-to-last line follows because the term $\mathit{recall}\cdot\log\left((1-\epsilon)/\epsilon\right)$ dominates all other terms except $(1-\epsilon)\log\epsilon$ which is constant. At this extreme, maximization of the cost function simply becomes maximization of recall, which is the same result as what we obtained for the cost function of stochastic neighbor embedding (in equation (3) in the main paper).

Special case 1. In the special case where $|P_i| = N-1$ and $|R_i| < N-1$, we set $a_i = 1/|P_i|$ and $b_i = 0$. From $|P_i| = N-1$ it follows that $N_{FP,i} = N_{TN,i} = 0$. The last two sums on the right-hand side of equation (3) are then left out, and equation (3) becomes

$$
\begin{aligned}
L(i) &= -\log(1+\gamma) + N_{TP,i}\cdot a_i\log(c_i+\gamma a_i) \\
&\quad + N_{MISS,i}\cdot a_i\log(d_i+\gamma a_i) \\
&= \mathit{const.} + N_{TP,i}\cdot a_i\log(c_i+\gamma a_i) \\
&\quad + N_{MISS,i}\cdot a_i\log(d_i+\gamma a_i) \\
&= -\log(1+\gamma) \\
&+ \frac{N_{TP,i}}{|P_i|}\log\left(\frac{\frac{|P_i|}{|R_i|}+\frac{\gamma}{1-\epsilon}}{\frac{\gamma}{1-\epsilon}+\frac{\epsilon}{1-\epsilon}\cdot\frac{|P_i|}{N-|R_i|-1}}\right) \\
&+ \log\left(\frac{\epsilon}{N-|R_i|-1}+\frac{\gamma}{|P_i|}\right) \qquad (10)
\end{aligned}
$$

where the right-hand side follows by rearranging terms. The right-hand side is nearly the same as the right-hand size of (6); when ϵ is much smaller than γ the above becomes

$$
\begin{aligned}
L(i) &\approx \mathit{const.} + \frac{N_{TP,i}}{|P_i|}\log\left(\frac{\frac{|P_i|}{|R_i|}+\gamma}{\gamma}\right) + \log\left(\frac{\gamma}{|P_i|}\right) \\
&= \mathit{const.} + \mathit{recall}\cdot\log\left(1+\frac{1}{\gamma}\cdot\frac{\mathit{precision}}{\mathit{recall}}\right) \qquad (11)
\end{aligned}
$$

which is the same result as in the general case.

Special case 2. In the special case where $|P_i| < N-1$ and $|R_i| = N-1$, We set $c_i = 1/|R_i|$ and $d_i = 0$. From $|R_i| = N-1$ it follows that $N_{MISS,i} = N_{TN,i} = 0$. The second and fourth sums on the right-hand side of equation (3) are then left out, and the equation becomes

$$
\begin{aligned}
L(i) &= -\log(1+\gamma) + N_{TP,i}\cdot a_i\log(c_i+\gamma a_i) \\
&\quad + N_{FP,i}\cdot b_i\log(c_i+\gamma b_i) \\
&= -\log(1+\gamma) + N_{TP,i}\frac{1-\epsilon}{|P_i|}\log\left(\frac{1}{|R_i|}+\gamma\frac{1-\epsilon}{|P_i|}\right) \\
&+ \frac{N_{FP,i}\cdot\epsilon}{N-|P_i|-1}\log\left(\frac{1}{|R_i|}+\gamma\frac{\epsilon}{N-|P_i|-1}\right) \qquad (12)
\end{aligned}
$$

which for small ϵ becomes

$$
\begin{aligned}
L(i) &\approx -\log(1+\gamma) + \frac{N_{TP,i}}{|P_i|}(1-\epsilon)\Bigg[\log(1-\epsilon) \\
&\quad + \log\left(\frac{1}{|R_i|}+\frac{\gamma}{|P_i|}\right)\Bigg] \\
&\approx -\log(1+\gamma) + \frac{N_{TP,i}}{|P_i|}\log\left(\frac{1}{|R_i|}+\frac{\gamma}{|P_i|}\right) \qquad (13)
\end{aligned}
$$

Since $|R_i| = N - 1$ we also have $N_{TP,i} = |P_i|$ and the above becomes

$$
\begin{aligned}
L(i) \approx & -\log(1+\gamma) + \frac{N_{TP,i}}{|P_i|} \log\left(\frac{1}{|R_i|} + \frac{\gamma}{|P_i|}\right) \\
& + \frac{N_{TP,i}}{|P_i|}\left(\log\left(\frac{\gamma}{|P_i|}\right) - \log\left(\frac{\gamma}{|P_i|}\right)\right) \\
= & -\log(1+\gamma) + \frac{N_{TP,i}}{|P_i|} \log\left(\frac{|P_i|}{|R_i|}\frac{1}{\gamma} + 1\right) \\
& + \frac{|P_i|}{|P_i|} \log\left(\frac{\gamma}{|P_i|}\right) \\
= & -\log(1+\gamma) + recall \cdot \log\left(\frac{precision}{recall}\frac{1}{\gamma} + 1\right) \\
& + \log\left(\frac{\gamma}{|P_i|}\right) \\
= & \; const. + recall \cdot \log\left(1 + \frac{1}{\gamma} \cdot \frac{precision}{recall}\right) \qquad (14)
\end{aligned}
$$

which is the same result as in the general case.

Special case 3. In the special case where $|P_i| = |R_i| = N - 1$, we set $a_i = 1/|P_i|$, $c_i = 1/|R_i|$ and $b_i = d_i = 0$. From $|P_i| = |R_i| = N - 1$ it follows that $N_{MISS,i} = N_{FP,i} = N_{TN,i} = 0$. The second, third, and fourth sums on the right-hand side of equation (3) are then left out, and we have

$$
\begin{aligned}
L(i) &= -\log(1+\gamma) + N_{TP,i} \cdot a_i \log(c_i + \gamma a_i) \\
&= -\log(1+\gamma) + \frac{N_{TP,i}}{|P_i|} \log\left(\frac{1}{|R_i|} + \frac{\gamma}{|P_i|}\right) \qquad (15)
\end{aligned}
$$

which is the same as the right-hand size of (13), and the analysis then proceeds in the same way as in Special case 2, again yielding the same result as in the general case.