vectorized informativity

Abstract

The purpose of this document is to explain how measuring informativity (informativeness, communicative success) of speakers and listeners can be a vectorized computation.

1 Preliminaries

A language L defined as a set of expressions E, each denoting a subset of the meaning space M. A speaker (sender) S and a listener (receiver) R are conditional distributions representing how the agents use L.

- $p_m \in \mathbb{R}^{|M|}$ is the prior over meanings
- $S \in \mathbb{R}^{|M| \times |E|}$ is the sender conditional probabilities of expressions given meanings.
- $R \in \mathbb{R}^{|E| \times |M|}$ is the receiver conditional probabilities of meanings given expressions.
- $\mathrm{diag}(P) \in \mathbb{R}^{|M| \times |M|}$ is the diagonal matrix of p.
- $U \in \mathbb{R}^{|M| \times |M|}$ is the utility matrix with $U_{ij} = u(m_i, m_j)$.

2 Literal Informativity

The **literal informativity** of a language L is the (joint) expected utility of speaker and listener behavior with respect to the meanings in L:

$$
\begin{aligned}
I(S; R) &= \mathbb{E}_P[u(m, m')] \\
&= \sum_{m,m'} P(m, m') \cdot u(m, m') \\
&= \sum_{m,m'} \left(P(m) \sum_{e} P(e|m) P(m'|e) \right) u(m, m')
\end{aligned}
$$

Consider the matrix product $Q = SR$, e.g.

$$
\begin{aligned}
&Q_{ij} \\
&= S_i^\top R_j \\
&= \begin{bmatrix} p(e_1|m_i) & \dots & p(e_{|E|}|m_i) \end{bmatrix} \begin{bmatrix} p(m_j|e_1) \\ \cdot \\ \cdot \\ \cdot \\ p(m_j|e_{|E|}) \end{bmatrix} \\
&= p(e_1|m_i)p(m_j|e_1) + \cdots + p(e_{|E|}|m_i)p(m_j|e_{|E|}) \\
&= \sum_e p(e|m_i)p(m_j|e)
\end{aligned}
$$

The joint probability of speaker and listener meanings is the matrix product $P_{m,m'} = \text{diag}(P)Q$ such that:

$$
\begin{aligned}
P(m_i, m_j) &= \text{diag}(P)_i Q_j \\
&= \begin{bmatrix} 0 & \dots & p(m_i) & \dots & 0 \end{bmatrix} \begin{bmatrix} \sum_e p(e|m_1)p(m_j|e) \\ \cdot \\ \cdot \\ \cdot \\ \sum_e p(e|m_j)p(m_j|e) \\ \cdot \\ \cdot \\ \cdot \\ \sum_e p(e|m_{|M|})p(m_j|e) \end{bmatrix} \\
&= p(m_i) \sum_e p(e|m_i)p(m_j|e)
\end{aligned}
$$

This allows us to rewrite the literal informativity as follows:

$$
\begin{aligned}
I(S; R) &= \sum_{m,m'} P(m, m') \cdot u(m, m') \\
&= \sum_{m,m'} \text{diag}(P) SR \odot U
\end{aligned}
$$

Where $\odot$ is element-wise multiplication.

3 Pragmatic Informativity

Pragmatic Listener

The pragmatic listener is a matrix $R^{(\text{prag})} \in \mathbb{R}^{|E| \times |M|}$ such that

$$R_{ij}^{(\text{prag})} = P_{\text{pragmatic listener}}(m_j|e_i) \propto P_{\text{pragmatic speaker}}(e_i|m_j) \cdot p(m_j)$$

Specifically,

$$P_{\text{pragmatic listener}}(m_j|e_i) = \frac{P_{\text{pragmatic speaker}}(e_i|m_j) \cdot p(m_j)}{\sum_{k \in |M|} P_{\text{pragmatic speaker}}(e_i|m_j) \cdot p(m_k)}$$

Pragmatic Speaker

The pragmatic speaker is a matrix $S^{(\text{prag})} \in \mathbb{R}^{|M| \times |E|}$ such that

$$S_{ij}^{(\text{prag})} = P_{\text{pragmatic speaker}}(e_i|m_j) \propto \exp(\alpha \cdot U^{(\text{pragmatic speaker})}(e_i, m_j))$$

Where

$$U^{(\text{pragmatic speaker})}(e_i, m_j) = \log P_{\text{literal listener}}(m_j|e_i)$$

$$= \log R_{ji}^{(\text{lit})}$$

And $\alpha \cdot$ is the *temperature* parameter defining the ‘optimality’ of the rationality of the pragmatic speaker; when no strong assumptions are made, it takes the value 1.0.

Literal Listener

The literal listener $R^{(\text{lit})}$ is just the matrix $R \in \mathbb{R}^{|E| \times |M|}$ defined in Section 2.

Vectorizing the pragmatic speaker

The probability of expression e_j given some intended meaning m_i is the entry:

$$S_{ij}^{(\text{prag})} = P_{\text{pragmatic listener}}(m_j|e_i)$$

$$= \frac{\exp(\alpha \cdot \log R_{ji}^{(\text{lit})})}{\sum_{k \in |E|} \exp(\alpha \cdot \log R_{ki}^{(\text{lit})})}$$

The probability distribution over expression choices given some intended meaning m_i is the row vector:

$$S_i^{(\text{prag})} = P_{\text{pragmatic speaker}}(e|m_i)$$

$$= \left[\frac{\exp(\alpha \cdot \log R_{1i}^{(\text{lit})})}{\sum_{k \in |E|} \exp(\alpha \cdot \log R_{ki}^{(\text{lit})})} \quad \cdots \quad \frac{\exp(\alpha \cdot \log R_{|E|i}^{(\text{lit})})}{\sum_{k \in |E|} \exp(\alpha \cdot \log R_{ki}^{(\text{lit})})} \right]$$

$$= \frac{\exp(\alpha \cdot \log R_{:,i}^{(\text{lit})})}{\sum_k \exp(\alpha \cdot \log R_{k,i}^{(\text{lit})})}$$

When $\alpha = 1.0,$ we have

$$= \text{softmax}(\log R_{:,i}^{(\text{lit})})$$

$$= R_{:,i}^{(\text{lit})}$$

Vectorizing the pragmatic listener

The probability of guessing a meaning j given the heard expression i is the entry:

$$\begin{aligned} R_{ij}^{(\text{prag})} &= P_{\text{pragmatic listener}}(m_j|e_i) \\ &= \frac{P_{\text{pragmatic speaker}}(e_i|m_j) \cdot p(m_j)}{\sum_{m \in |M|} P_{\text{pragmatic speaker}}(e_i|m_k) \cdot p(m_k)} \end{aligned}$$

The probability distribution over meaning guesses given the heard expression i is the row vector:

$$\begin{aligned} R_i^{(\text{prag})} &= P_{\text{pragmatic listener}}(m|e_i) \\ &= \left[\frac{P_{\text{pragmatic speaker}}(e_i|m_1) \cdot p(m_1)}{\sum_{k \in |M|} P_{\text{pragmatic speaker}}(e_i|m_k) \cdot p(m_k)} \quad \cdots \quad \frac{P_{\text{pragmatic speaker}}(e_i|m_{|M|}) \cdot p(m_{|M|})}{\sum_{k \in |M|} P_{\text{pragmatic speaker}}(e_i|m_k) \cdot p(m_k)} \right] \\ &= \left[\frac{S_{1i}^{(\text{prag})} \cdot p(m_1)}{\sum_{k \in |M|} S_{ki}^{(\text{prag})} \cdot p(m_k)} \quad \cdots \quad \frac{S_{|M|i}^{(\text{prag})} \cdot p(m_{|M|})}{\sum_{k \in |M|} S_{ki}^{(\text{prag})} \cdot p(m_k)} \right] \\ &= \frac{S_{:,i}^{(\text{prag})\top} p_m}{\sum_k S_{k,i}^{(\text{prag})\top} p_m} \end{aligned}$$

3.1 Informativity

The **pragmatic informativity** of a language L is

$$\begin{aligned} I(S^{(\text{prag})}; R^{(\text{prag})}) &= \mathbb{E}_P[u(m, m')] \\ &= \sum_{m, m' \in M} P(m) \sum_{e \in L} P_{\text{pragmatic speaker}}(e|m) P_{\text{pragmatic listener}}(m'|e) \cdot u(m, m') \\ &= \sum_{m, m' \in M} \text{diag}(P) S^{(\text{prag})} R^{(\text{prag})} \odot U \end{aligned}$$