Modelling in Neuroscience

Agglomerative Information Bottleneck

Nicolás Violante

Introduction

The Information Bottleneck

$$\min_{p(\tilde{x}|x)} I(X, \tilde{X}) - \beta I(\tilde{X}, Y)$$

- Find compact representation of X
- Retain as much meaningful information as possible with respect to Y

The Information Bottleneck

- Set of self-consistent equations

$$\begin{cases} p(\tilde{x}|x) = \frac{1}{Z(\beta,x)} p(\tilde{x}) e^{-\beta \mathrm{KL}(p(y|x)||p(y|\tilde{x}))} \\ p(\tilde{x}) = \sum_x p(\tilde{x}|x) p(x) \\ p(y|\tilde{x}) = \sum_x p(\tilde{x}|x) p(y|x) \frac{p(x)}{p(\tilde{x})} \end{cases}$$

The Hard Clustering Limit

- Limit case with $\beta \to \infty$

- The equations become

$$\begin{cases} p(\tilde{x}|x) = \begin{cases} 1 \text{ if } x \in \tilde{x} \\ 0 \text{ otherwise} \end{cases} \\ p(\tilde{x}) = \sum_x p(\tilde{x}, x) \\ p(y|\tilde{x}) = \sum_x p(\tilde{x}|x) p(y|x) \frac{p(x)}{p(\tilde{x})} \end{cases}$$

Agglomerative Information Bottleneck

- Idea: given a partition, merge its two elements that yield minimal decrease of mutual information

How to evaluate the best merge?

- Y-decrease information

$$\delta I_Y(z_i, z_j) = I(Z_m, Y) - I(Z_{m-1}, Y)$$

$$\delta I_Y(z_i, z_j) = (p(z_i) + p(z_j))\mathrm{JS}(p(y|z_i), p(y|z_j))$$

Algorithm

Input: Joint probability distribution $p(x, y)$

Output: A partition of X into m clusters, $\forall m \in \{1...|X|\}$

Initialization:

- Construct $\tilde{X} \equiv X$
- $\forall i, j = 1...|X|,\ i < j$, calculate
  $d_{i,j} = (p(\tilde{x}_i) + p(\tilde{x}_j)) D_{JS}[p(y|\tilde{x}_i), p(y|\tilde{x}_j)]$

Loop:

- For $m = |X| - 1 \ ... \ 1$
  - Find the indices $\{i, j\}$ for which $d_{i,j}$ is minimized
  - Merge $\{\tilde{x}_i, \tilde{x}_j\} \Rightarrow \tilde{x}_*$
  - Update $\tilde{X} = \{\tilde{X} - \{\tilde{x}_i, \tilde{x}_j\}\} \bigcup \{\tilde{x}_*\}$
  - Update $d_{i,j}$ costs w.r.t. $\tilde{x}_*$
- **End For**

Experiments

Simple example

	Y=0	**Y=1**
X=0	0.16	0.04
X=1	0.255	0.045
X=2	0.27	0.03
X=3	0.02	0.08
X=4	0.03	0.07

$$Z_5 = \{\{0\}, \{1\}, \{2\}, \{3\}, \{4\}\}$$

$$Z_4 = \{\{2\}, \{3\}, \{4\}, \{0, 1\}\}$$

$$Z_3 = \{\{2\}, \{3\}, \{4, 0, 1\}\}$$

$$Z_2 = \{\{2, 3\}, \{4, 0, 1\}\}$$

$$Z_1 = \{0, 1, 2, 3, 4\}$$

Word Clustering

Dataset

- 20NewsGroup dataset
- 1500 online posts about science and politics

Data pre-processing

- Remove meta-data
- Remove punctuation signs and lower case
- Lemmatization
- Keep nouns only

Some clusters

['information'], ['year'], ['chip'], ['number'], ['fact'], ['people'], ['right'], ['country'], ['time'], ['encryption'], ['weapon'], ['government'], ['thing'], ['system'], ['datum'], ['state'], ['space']
['information'], ['chip'], ['fact'], ['right'], ['encryption'], ['weapon'], ['government'], ['space'], ['year', 'thing'], ['people', 'time'], ['number', 'system', 'datum', 'country', 'state']
['chip'], ['right'], ['government'], ['space'], ['year', 'thing'], ['fact', 'people', 'time'], ['encryption', 'weapon', 'information', 'number', 'system', 'datum', 'country', 'state']
['space'], ['government', 'encryption', 'weapon', 'information', 'number', 'system', 'datum', 'country', 'state', 'chip', 'year', 'thing'], ['right', 'fact', 'people', 'time']

Evaluation

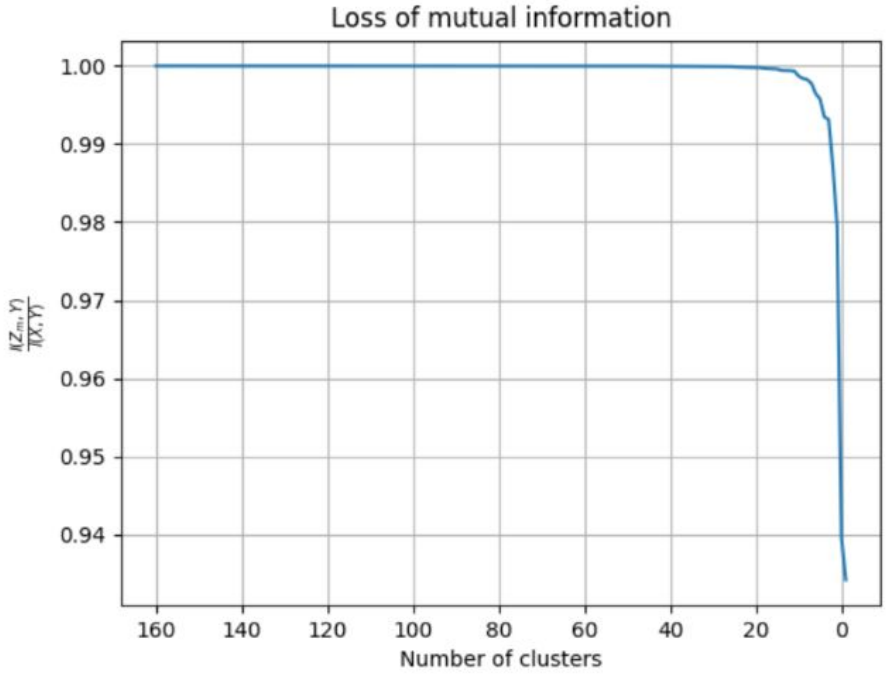

Figure 2: Evolution of the loss of information $I(Z_m, Y)$ due to the clustering

Thank you