Intelligent Systems: Reasoning and Recognition

James L. Crowley

ENSIMAG 2 / MoSIG M1 Second Semester 2012/2013

Lesson 13 - Exercise 5 3 April 2013
(updated 21 may 2013)

Exercise : Bayesian Reasoning

Bayesian Reasoning is a widely used technique to validate or invalidate hypothesis using uncertain or unreliable information. With this approach, a set of K possible hypotheses, H_k, are formulated and assigned a probability, $P(H_k)$. As new evidence, E, for or against the hypothesis is obtained it is also assigned a probability $P(E)$ as well as a probability that it confirms the hypothesis, $P(E|H_k)$. Baye's rule is then used to update the probability of each of the hypotheses:

$$P(H_k \mid E) = \frac{P(E \mid H_k)P(H_k)}{\sum_{j=1}^{K} P(E \mid H_j)P(H_j)}$$

This rule is applied recursively as new evidence is obtained. Let us define $S=\{E_n\}$ as a body of previous evidence composed of N elements, and E as a new element of evidence. Then Baye's rule tells us that:

$$P(H_k \mid E,S) \leftarrow \frac{P(E,S \mid H_k)P(H_k)}{\sum_{j=1}^{K} P(E,S \mid H_j)P(H_j)}$$

However, this poses the problem of how to represent the joint probability of the new evidence, E, and all of the past evidence S. We will solve this by assuming conditional independence. While this is not strictly true, it is a reasonable hypothesis and provides a reasonable answer. In this case assuming E and S are conditionally independent, allows us to apply the chain rule:

$$P(E,S \mid H_k) = P(E \mid H_k)P(S \mid H_k)$$

Which gives $$P(H_k \mid E,S) \leftarrow \frac{P(E \mid H_k)P(S \mid H_k)P(H_k)}{\sum_{j=1}^{K} P(E \mid H_j)P(S \mid H_j)P(H_j)}$$

Note that the values for the cumulative evidence $P(S \mid H_k)$ are in fact a product of probabilities the N individual evidences, E_n.

$$P(H_k \mid S) = \frac{\prod_{n=1}^{N} P(E_n \mid H_k)P(H_k)}{\sum_{j=1}^{K}\prod_{n=1}^{N} P(E_n \mid H_j)P(H_j)}$$

However, we accumulate this product recursively as we update the probabilities of each hypothesis with each new evidence.

In this exercise we will use this approach to classify a body of text based on the frequency of occurrence of words. This can be used by a text editor to recognize the category of document that a person is composing (for example personal letters, technical reports, or computer code) and propose appropriate formatting, grammar and spelling corrections. Histograms (or bags) of words can be used to estimate the required probabilities for $P(E|H)$ and $P(E)$. For this task, assume that you have a training corpus composed of $K=5$ classes of text, and that for each class you have a sample composed of M words.

1) Explain how the training corpus can be used to construct a table for the frequency of occurrence for each word in the training data for each class of text. What it the probability that word, W, will occur in a sample from Class K? What is the probability that W will occur anywhere in the corpus?

2) Propose a method to obtain an initial estimate for the probability that an unknown text (a probe) belongs to each class.

3) Explain how to update the estimate for the probability of each class as the user types each new word in the probe (the unknown text).

4) What happens if the probe contains a word that was not in the training corpus? What can you do to protect against this case? How do you update the class estimates?

5) Is it necessary to recompute the bags of words if the user decides to create a new class of document, and provides a new corpus for this class?