Test

The verification of each example is a process that requires classifying a generated example in the teacher's preferred class of examples. The system keeps records of the examples that each teacher uses or has used most frequently, as explained previously in this section. The classification of an example implies calculating the probability of the example *a* belonging to the class of examples that the teacher prefers. Given

$$a \in A$$

where

a is an example

A is the set of all the acceptable examples

ARIALE uses a naïve Bayesian classifier (see also section 2.2.7.) to calculate this probability. The classifier is intended to estimate $P(ai|class)$, the probability of a specific example a_i, to belong to a particular class of topologies. As this calculation is difficult, thus we use the probability of a specific class given the evidence a_i.

$$P(class|a_i) = P(class)\, P(a_i|class) \tag{16}$$

where

class is the class of examples for which the generator is testing to see if the current example belongs to it.

a_i is the evidence

A network topology has four indices, and any example a_i can also be represented as a vector of four indices:

$a_i = (r, m, b, c)$

where

r is the network reliability

m is the number of links

b is the total network capacity

c is the network cost

For example, a given topology a_1 can be represented as a vector of its measures of performance and number of links. One example could be a network with the following values for each measure of performance:

Reliability = 2

Number of links = 11

Capacity = 21214

Cost = 144350

Then, the representation of this example is the vector

a_1 = (2,11,21214,144350)

and we have to check in which of the classes described in section 5.2 this example can be classified. According to Table 5.5., we are representing the capacity and cost values with nominal values such as “low”, “medium” and “high”. Because the system basically requires checking if a topology belongs to a particular class or not, we are using here a two-class classification to explain how the Bayesian classification classifies topologies. The two classes are “A” and “B” and the example to be classified is the one detailed previously. For this explanation we use the following training dataset shown in Table 5.9.:

Table 5.9. A training dataset.

Id	*Reliability (R)*	*Number of links (L)*	*Capacity (Ca)*	*Costo (Co)*	*Class*
1	2	7	High	High	A
2	2	16	High	High	A
3	2	9	Medium	High	A
4	2	14	Medium	High	A
5	2	11	Medium	Low	A
6	2	12	Medium	Low	A
7	2	10	Medium	Medium	A
8	2	13	Medium	Medium	A
9	3	15	High	High	A
10	3	17	Low	Low	A
11	3	18	Low	Low	A
12	3	15	Medium	High	A
13	3	16	Medium	High	A
14	3	11	Medium	Low	A
15	3	13	Medium	Low	A
16	3	17	Medium	Medium	A
17	5	19	High	Medium	B
18	4	18	High	High	B
19	5	19	High	High	B
20	5	18	Medium	High	B
21	4	18	Medium	High	B
22	4	18	Medium	High	B
23	4	19	Medium	High	B
24	4	19	High	Medium	B
25	5	18	Medium	High	B
26	4	19	Medium	High	B
27	4	19	High	High	B
28	4	19	High	High	B
29	4	18	High	High	B
30	5	19	High	Medium	B
31	4	19	High	High	B
32	4	19	Medium	High	B
33	4	17	High	High	B
34	5	20	High	High	B
35	4	18	Medium	High	B

We must calculate the probabilities for each measure of performance. In order to make the comprehension of the calculations easier, we use the following letters between parentheses to identify each class and measure of performance:

Class A (A)

Class B (B)

Reliability (Re)

Number of links (Li)

Capacity (Ca)

Cost (Co)

Given the example to be classified and the training dataset, for the measure of performance, for example in the case of *cost*, it is necessary to calculate its different probabilities of appearing in each class. Formally,

$$P(Co|class) \tag{17}$$

where

Co is the value of the measure of performance Cost

class is the class of examples for which the generator is testing if the measure of performance is included

$P(\mathrm{Co}|class)$ is the probability that *class* presents *Co* as evidence.

For each value in each position of the vector, the naïve Bayesian classifier uses equation (11) to calculate its probability of belonging to each class of examples. Equation (18) shows equation (11) adapted to calculate the probabilities of types of link

$$P(f_i|v_j) = \frac{n_c + mp}{n+m} \tag{18}$$

where

f_j is each value of each measure of performance (Re, Li, Ca, Co) in the vector for each class,

v_j is each class

n is the total number of instances in each class v_j

n_c is the number of instances with attribute f_i and class v_j

p is a priori estimate of $P(f_i|v_j)$ for each f_j. We assume that the probabilities of all attributes are equiprobable (equally likely to be true). For example, for the attribute “reliability” there are four possible values and p = 1/4.

m is a constant used to avoid the possible consequences that could arise if $n_c = 0$ (in this case the calculation would be zero). We use m = 4 as a constant because we have four attributes in the examples of our training dataset.

This equation is applied to calculate the probability that each class include the given topology. Table 5.10. shows the corresponding values for the calculation of probabilities using equation (18).

Table 5.10. Values to calculate where the topology can be classified.

Class A	**Class B**
Re (2) $n = 16$ $nc = 8$ $p = .25$ $m = 4$	Re (2) $n = 19$ $nc = 0$ $p = .25$ $m = 4$
Li (11) $n = 16$ $nc = 11$ $p = .08$ $m = 2$	Li (11) $n = 19$ $nc = 8$ $p = .08$ $m = 2$
Ca (Medium) $n = 16$ $nc = 3$ $p = .33$ $m = 4$	Ca (Medium) $n = 19$ $nc = 3$ $p = .33$ $m = 4$
Co (Medium) $n = 16$ $nc = 2$ $p = .33$ $m = 4$	Co (Medium) $n = 19$ $nc = 0$ $p = .33$ $m = 4$

Table 5.11. shows the corresponding calculation of the probabilities of each measure of performance of the given example for each class.

Table 5.11. Probabilities for each attribute for each class.

Class A	**Class B**
$P(\mathrm{Re}\|A)=\frac{8+4*.25}{16+4}=0.2$	$P(\mathrm{Re}\ B)=\frac{0+4*.25}{19+4}=0.04$
$P(Li\|A)=\frac{2+4*.07}{16+4}=0.12$	$P(Li\|B)=\frac{0+4*.07}{19+4}=0.01$
$P(Ca\|A)=\frac{11+4*.33}{16+4}=0.62$	$P(Ca\|B)=\frac{8+4*.33}{19+4}=0.04$
$P(Co\|A)=\frac{3+4*.33}{16+4}=0.22$	$P(Co\|B)=\frac{3+4*.33}{19+4}=0.19$

After these calculations, it is possible to multiply the probabilities of each measure of performance to appear in a class in order to obtain the global probability of the vector to belong to a particular class. The maximal probability obtained determines to which class the topology can belong, according to equation (10) in section 2.2.7., which we adjust as follows for the current example

$$P(v_j)\prod_i P(a_i|v_j) \tag{19}$$

where

{A,B } is the set of the different classes to classify examples, and

E is the example to be classified.

This calculation can also be performed as follows:

P(A)*P(RE/A)*P(Li/A)* P(Ca/A)* P(Co/A) = 0,472 * 0.2 * 0.12 * 0.62 * 0.22 = 0,001456019

P(B)*P(Re/B)*P(Li/B) *P(Ca/B) *P(Co/B) = 0,527 * 0.04 * 0.01 * 0.41 * 0.19 = 0,000023470

Finally, the classifier has a prediction for each class. The maximal number identifies the class to classify our vector *a*. In this example, the class of the given example is “A” because 0,001456019 > 0,000023470.