Exercises in Bayesian reasoning: Proceedings of the Ministry of Magic

Joachim Vandekerckhove

The happy herbologist

The happy herbologist

At Hogwarts School of Witchcraft and Wizardry, professor Pomona Sprout leads the Herbology Department. In one project, she is developing a spell to evaluate the quality of a magical plant called *green codacle*—a leaf that when consumed causes a witch or wizard to feel euphoric and relaxed. However, a mutation affecting one in a thousand codacle plants changes its effects, and those rare plants cause paranoia and anxiety.

The happy herbologist

At Hogwarts School of Witchcraft and Wizardry, professor Pomona Sprout leads the Herbology Department. In one project, she is developing a spell to evaluate the quality of a magical plant called *green codacle*—a leaf that when consumed causes a witch or wizard to feel euphoric and relaxed. However, a mutation affecting one in a thousand codacle plants changes its effects, and those rare plants cause paranoia and anxiety.

Professor Sprout's new spell can accurately detect an existing mutation 99% of the time, but will also falsely indicate that a healthy plant is a mutant 2% of the time.

The happy herbologist

At Hogwarts School of Witchcraft and Wizardry, professor Pomona Sprout leads the Herbology Department. In one project, she is developing a spell to evaluate the quality of a magical plant called *green codacle*—a leaf that when consumed causes a witch or wizard to feel euphoric and relaxed. However, a mutation affecting one in a thousand codacle plants changes its effects, and those rare plants cause paranoia and anxiety.

Professor Sprout's new spell can accurately detect an existing mutation 99% of the time, but will also falsely indicate that a healthy plant is a mutant 2% of the time.

When Sprout presents her results at a School colloquium, Professor of Divination Sybill Trelawney asks two questions:

The happy herbologist

At Hogwarts School of Witchcraft and Wizardry, professor Pomona Sprout leads the Herbology Department. In one project, she is developing a spell to evaluate the quality of a magical plant called *green codacle*—a leaf that when consumed causes a witch or wizard to feel euphoric and relaxed. However, a mutation affecting one in a thousand codacle plants changes its effects, and those rare plants cause paranoia and anxiety.

Professor Sprout's new spell can accurately detect an existing mutation 99% of the time, but will also falsely indicate that a healthy plant is a mutant 2% of the time.

When Sprout presents her results at a School colloquium, Professor of Divination Sybill Trelawney asks two questions:

- What is the probability that a codacle plant is a mutant, when your spell says that it is?
- And what is the probability the plant is a mutant, when your spell says that it is healthy?

The happy herbologist

Call the event that a specific plant is a mutant $\mathcal{M}$, and that it is healthy $\neg\mathcal{M}$.

Call the event that professor Sprout's spell diagnoses a plant as a mutant D, and that it diagnoses it healthy $\neg D$.

Professor Trelawney's interest is in the probability the plant is indeed a mutant given it has been diagnosed as a mutant, or $P(\mathcal{M}|D)$, and the probability the plant is a mutant given it has been diagnosed healthy, or $P(\mathcal{M}|\neg D)$.

The happy herbologist

Professor Trelawney, who is an accomplished statistician, has all the relevant information to apply Bayes' Rule to find these probabilities:

- She knows the prior probability that a plant is a mutant is $P(\mathcal{M}) = .001$, and thus the prior probability that a plant is not a mutant is $P(\neg\mathcal{M}) = 1 - P(\mathcal{M}) = .999$.
- The probability of a correct mutant diagnosis given the plant is a mutant is $P(D|\mathcal{M}) = .99$, and the probability of an erroneous healthy diagnosis given the plant is a mutant is thus $P(\neg D|\mathcal{M}) = 1 - P(D|\mathcal{M}) = .01$.
- When the plant is healthy, the spell incorrectly diagnoses it as a mutant with probability $P(D|\neg\mathcal{M}) = .02$, and correctly diagnoses the plant as healthy with probability $P(\neg D|\neg\mathcal{M}) = 1 - P(D|\neg\mathcal{M}) = .98$.

The happy herbologist

When Professor Sprout's spell gives a mutant diagnosis, the posterior probability that the plant is really a mutant is given by:

$$P(\mathcal{M}|D) = \frac{P(\mathcal{M})P(D|\mathcal{M})}{P(\mathcal{M})P(D|\mathcal{M}) + P(\neg\mathcal{M})P(D|\neg\mathcal{M})}.$$

Professor Trelawney can now plug in the values to find that the posterior probability the plant is a mutant given a mutant diagnosis is:

$$P(\mathcal{M}|D) = \frac{.001 \times .99}{.001 \times .99 + .999 \times .02} \approx .047.$$

The happy herbologist

A mutant diagnosis from Professor Sprout's spell raises the probability the plant is a mutant from .001 to roughly .047. This means that when a plant is diagnosed as a mutant, the posterior probability the plant is *not* a mutant is $P(\neg\mathcal{M}|D) = 1 - .047 \approx .953$.

The happy herbologist

A mutant diagnosis from Professor Sprout's spell raises the probability the plant is a mutant from .001 to roughly .047. This means that when a plant is diagnosed as a mutant, the posterior probability the plant is *not* a mutant is $P(\neg\mathcal{M}|D) = 1 - .047 \approx .953$.

The low prior probability that a plant is a mutant means that, even with the spell having 99% accuracy to correctly diagnose a mutant plant as such, a plant diagnosed as a mutant is still probably safe to eat – but Professor Trelawney might now think twice.

The happy herbologist

Analogous calculations show that the posterior probability that a plant is a dangerous mutant, given it is diagnosed as healthy, is:

$$P(\mathcal{M}|\neg D) = \frac{.001 \times .01}{.001 \times .01 + .999 \times .98} \approx .000010.$$

The happy herbologist

Analogous calculations show that the posterior probability that a plant is a dangerous mutant, given it is diagnosed as healthy, is:

$$P(\mathcal{M}|\neg D) = \frac{.001 \times .01}{.001 \times .01 + .999 \times .98} \approx .000010.$$

The posterior probability that a plant is a dangerous mutant despite being diagnosed as healthy is almost vanishingly small, so Professor Trelawney can be relatively confident she is eating a healthy plant after professor Sprout's spell returns a healthy diagnosis.

The happy herbologist

An advantage of using Bayes' Rule in this way is that it gracefully extends to more complex scenarios.

The happy herbologist

An advantage of using Bayes' Rule in this way is that it gracefully extends to more complex scenarios.

Consider the perhaps disappointing value of $P(\mathcal{M}|D)$: a mutant diagnosis only brings the posterior probability up to 5%. Suppose, however, that Professor Sprout knows that her diagnosis is statistically independent from the diagnosis of her research associate Neville Longbottom, and suppose that both Sprout and Longbottom return the mutant diagnosis. Due to the independence of the diagnoses, we know that

$$\begin{aligned} P(D_S, D_L|\mathcal{M}) &= P(D_S|\mathcal{M}) \times P(D_L|\mathcal{M}) = .99 \times .99 = .9801 \\ P(D_S, D_L|\neg\mathcal{M}) &= P(D_S|\neg\mathcal{M}) \times P(D_L|\neg\mathcal{M}) = .02 \times .02 = .0004 \end{aligned}$$

The happy herbologist

Applying Bayes' Rule once more (and considering that the joint event (D_S, D_L) can be treated like any other event), the posterior probability that the plant is really a mutant is now given by:

$$\begin{aligned} P(\mathcal{M}|D_S, D_L) &= \frac{P(\mathcal{M})P(D_S, D_L|\mathcal{M})}{P(\mathcal{M})P(D_S, D_L|\mathcal{M}) + P(\neg\mathcal{M})P(D_S, D_L|\neg\mathcal{M})} \\ &= \frac{.001 \times .99 \times .99}{.001 \times .99 \times .99 + .999 \times .02 \times .02} \approx .71, \end{aligned}$$

which illustrates the value of multiple independent sources of evidence: a plant that has been independently diagnosed as a mutant is quite likely to be one. A third independent diagnosis would put the posterior probability over 99%.