A METHOD FOR COMBINING JUDGEMENTS IN DISTRIBUTED DECISION MAKING, APPLIED TO THE STARDUST PROJECT. J. S. Von Korff, A. J. Westphal, D. P. Anderson, *Space Sciences Laboratory, University of California at Berkeley, Berkeley, CA 94720-7450, USA, (vonkorff@socrates.berkeley.edu).*

1 Introduction

In January 2006, the STARDUST mission will return samples of cometary, interplanetary, and interstellar dust captured in aerogel collectors. In order to locate the interstellar dust grains, an automated microscope will scan images of the aerogel. We expect to scan 1.5 million ''focus movies,'' where each ''focus movie'' consists of about 40 images from the same horizontal location in the aerogel, but at different focus depths. However, we expect only 50 interstellar dust grains in all. In order to search the scanned images for dust, we will deliver the images to volunteers via the internet. The volunteers will examine the movies, and will attempt to determine which of the movies contain interstellar dust tracks.

Some volunteers are bound to get the wrong answer, though. Either they will believe they see tracks where there are none, or they will miss seeing tracks that are actually present. Suppose the average volunteer, when presented with a movie that contains no track, is 5% likely to mistakenly record a track. This will result in 75,000 false positives, which is more than we would like. To avoid this problem, we should assign multiple volunteers to each movie. This abstract considers the method by which we determine how many volunteers to assign. The underlying mathematics is simple, and essentially dates back to [1].

2 Algorithm assuming independence of judges and perfect measurement of skill level

We will start with two simplifying assumptions, and consider later what happens if we relax the assumptions.

Assumption number one: First, call a movie ''positive'' if it contains a particle and ''negative'' otherwise. Then define two skill levels for each volunteer: the sensitivity, p_{yes}, and the specificity, p_{no}. We define p_{yes} to be the fraction of positive movies that the volunteer correctly labels as positive, and p_{no} to be the fraction of negative movies that the volunteer correctly labels as negative. Now, consider k volunteers $v_1 \ldots v_k$, and take any $m \leq k$. Then the fraction of positive movies which are answered correctly by volunteers $v_1, v_2, \ldots v_m$ and incorrectly by volunteers $v_{m+1}, v_{m+2}, \ldots v_k$, is equal to:

$$\prod_{1 \leq i \leq m} p_{\text{yes}_i} \times \prod_{m+1 \leq j \leq k} (1 - p_{\text{yes}_j}) \qquad (1)$$

with the analogous statement for negative movies. That is to say, the volunteers' votes, on a randomly selected positive (or negative) movie, are statistically independent.

Assumption number two: We can measure perfectly the skill levels of each volunteer, by measuring their performance on a finite set of test movies. For example, if we feed 5 positive test movies to a volunteer, and the volunteer approves 4 of the 5 movies, we will assume that we have accurately measured the volunteer's skill level p_{yes}, and that the volunteer will get 80% of all positive movies right.

Algorithm:

Each movie starts with a ''certainty ratio'' of $r = 1$. We select volunteers to judge this movie in any order we like. Whenever a volunteer judges the movie to be positive, we multiply the ratio r by the value $\frac{p_{\text{yes}}}{1-p_{\text{no}}}$ for that volunteer. Whenever a volunteer judges the movie to be negative, we multiply the ratio r by the value $\frac{1-p_{\text{yes}}}{p_{\text{no}}}$ for that volunteer. If we ever obtain $r > r_{\text{yes}}$ for some suitably chosen constant $r_{\text{yes}} >> 1$, we judge the movie to be probably positive, and pass it on to our researchers for verification. If we ever obtain $r < r_{\text{no}}$ for some suitably chosen constant $r_{\text{no}} << 1$, we judge the movie to be probably negative, and stop examining it.

Justification for the algorithm:

The algorithm is easily justified with a bit of Bayesian reasoning. Suppose we feed a given movie to some volunteers, and get a string of yes/no responses, $s_1, s_2, \ldots s_k$. Then there are two interpretations of the data. The first is that the movie is positive, in which case the probability of our data having occurred was

$$q_T = p \times \prod_{i|s_i=Y} p_{\text{yes}_i} \times \prod_{j|s_j=N} (1 - p_{\text{yes}_j}) \qquad (2)$$

where p is the fraction of movies that are positive. And the second is that the movie is negative, in which case the probability was

$$q_F = (1-p) \times \prod_{i|s_i=N} p_{\text{no}_i} \times \prod_{j|s_j=Y} (1 - p_{\text{no}_j}) \qquad (3)$$

Then the probability that the movie is positive is just

$$p_T = q_T/(q_T + q_F) = 1/(1 + q_F/q_T) = 1/(1 + \frac{1}{r}\frac{1-p}{p}) \qquad (4)$$

where r is the certainty ratio defined above. The larger the value of r, the more likely it is that the movie is positive. And the smaller the value of r, the more likely that the movie is negative. We need only wait until r reaches some suitably large or small value, in order to say that we are confident about our result. In particular, we can solve the above equation for r to find out the necessary r value for a desired p_T:

$$r_{\text{yes}} = \frac{p_T(1-p)}{(1-p_T)p} \qquad (5)$$

An analogous formula relates r_{no} to p_F.

3 Relaxing the assumptions

Next, let's talk about relaxing the assumptions above. Assumption number one is certainly faulty in practice, because we expect that some movies will be ''harder'' (meaning that a larger fraction of volunteers get it wrong) and others will be ''easier.'' A movie might be hard if it contains a speck that looks sort of like a particle track, but isn't. But if Eq. (1) holds, then almost all movies must be of approximately average difficulty; they cannot be hard or easy. (It's analogous to the situation in statistical mechanics, where almost all ensembles have energy near the average value.) Note that relaxing the independence assumption does not mean that we expect the volunteers to literally collude with one another; our setup does not allow that to happen.

So we would like to be able relax the first assumption; but it's hard to do so if we want to be able to say anything about the mathematics. Furthermore, we cannot hope for a completely general solution, because if Eq. (1) fails badly enough, it is possible that no algorithm can succeed. Consider what happens if the volunteers all have the same $p_{\rm yes}$ and $p_{\rm no}$, and all agree on every movie. Our task will be impossible, because there will be a fraction $(1 - p_{\rm yes})$ of the positive movies such that all volunteers believe them to be negative. So we cannot learn the true nature of these movies, no matter how many volunteers we ask.

The simplest heuristic, and the one we will adopt, is to hope that we can solve the problem by making the critical r values more extreme. If we carry out the algorithm and end up with too many false positives, we simply raise the value of $r_{\rm yes}$. (We can estimate the number of false positives since we know to expect about 50 positive movies.)

Next, let's consider relaxing assumption two. It is certainly false in practice that we can measure volunteers' skill levels exactly with any finite number of test movies. Instead, we should assume that the volunteers' skill levels are distributed according to some random variables. Let's suppose that we can find the function $F(p_{\rm yes}, p_{\rm no})$ which defines the joint distribution of $p_{\rm yes}$ and $p_{\rm no}$ over the set of all volunteers.

Then we apply Bayesian reasoning as before, but this time we include the number of test questions the volunteers answered correctly. Suppose we feed a given movie to some volunteers, and get a string of yes/no responses, $s_1, s_2, \ldots s_k$. And suppose the volunteers got $n_{{\rm yes}_1}, n_{{\rm yes}_2}, \ldots n_{{\rm yes}_k}$ of N test questions right when the answer was ''positive'', and $n_{{\rm no}_1}, n_{{\rm no}_2}, \ldots n_{{\rm no}_k}$ of N test questions right when the answer was ''negative.'' Then the probability of getting our data and the movie being positive is

$$q_T = p \times \int_{p_{{\rm yes}_m}, p_{{\rm no}_m} \text{ for all } m} \left(\prod_{\ell | s_\ell = Y} p_{{\rm yes}_\ell} \times \prod_{j | s_j = N} (1 - p_{{\rm yes}_j}) \right.$$
$$\times \prod_i F(p_{{\rm yes}_i}, p_{{\rm no}_i}) \binom{N}{n_{{\rm yes}_i}} {p_{{\rm yes}_i}}^{n_{{\rm yes}_i}} (1 - p_{{\rm yes}_i})^{N - n_{{\rm yes}_i}}$$
$$\left. \times \binom{N}{n_{{\rm no}_i}} {p_{{\rm no}_i}}^{n_{{\rm no}_i}} (1 - p_{{\rm no}_i})^{N - n_{{\rm no}_i}} \right) \quad (6)$$

To simplify this equation, let's define

$$< p_{{\rm yes}_i} > \equiv \int_{p_{{\rm yes}_i}, p_{{\rm no}_i}} p_{{\rm yes}_i} F(p_{{\rm yes}_i}, p_{{\rm no}_i})$$
$$\times \binom{N}{n_{{\rm yes}_i}} {p_{{\rm yes}_i}}^{n_{{\rm yes}_i}} (1 - p_{{\rm yes}_i})^{N - n_{{\rm yes}_i}}$$
$$\times \binom{N}{n_{{\rm no}_i}} {p_{{\rm no}_i}}^{n_{{\rm no}_i}} (1 - p_{{\rm no}_i})^{N - n_{{\rm no}_i}} \quad (7)$$

Note that this quantity depends only on the numbers $n_{{\rm yes}_i}$ and $n_{{\rm no}_i}$. Similarly, define

$$< 1 - p_{{\rm yes}_i} > \equiv \int_{p_{{\rm yes}_i}, p_{{\rm no}_i}} (1 - p_{{\rm yes}_i}) F(p_{{\rm yes}_i}, p_{{\rm no}_i}) \binom{N}{n_{{\rm yes}_i}} \ldots \quad (8)$$

Which also depends only on $n_{{\rm yes}_i}$ and $n_{{\rm no}_i}$. Then

$$q_T = p \times \prod_{i | s_i = Y} < p_{{\rm yes}_i} > \times \prod_{j | s_j = N} < 1 - p_{{\rm yes}_j} > \quad (9)$$

We get a similar equation for q_F. So our situation is the same as before, except that we must use $< p_{{\rm yes}_i} >, < 1 - p_{{\rm yes}_i} > \ldots$ in order to calculate the ratio r, whereas before we would have used $p_{{\rm yes}_i}, 1 - p_{{\rm yes}_i}, \ldots$. Calculating the values in angle brackets may seem to require performing complicated integrals each time a new volunteer comes along. However, since these values depend only on $n_{{\rm yes}_i}$ and $n_{{\rm no}_i}$, each of which has the range $0 \ldots N$, we need only compute $(N + 1)^2$ integrals total.

Thanks:

JVK would like to thank Jim Pitman for a helpful conversation.

References

[1] Poisson, S.-D. (1837) Recherches sur la probabilité des jugements