Asymmetric Information Measures: How to Extract Knowledge From an Expert so That the Expert's Effort Is Minimal

Hung T. Nguyen
Department of Mathematical Sciences
New Mexico State University
Las Cruces, New Mexico 88003, USA
Email: hunguyen@nmsu.edu

Vladik Kreinovich and Elizabeth Kamoroff
Department of Computer Science,
University of Texas at El Paso,
El Paso, TX 79968, USA
vladik@utep.edu

1. How Knowledge Is Extracted Now

- *Knowledge acquisition:* we ask experts questions, and put the answers into the computer system.
- *Problem:* it is a very time-consuming and therefore expensive task.
- *Objective:* minimize the effort of an expert.
- *Related problem:* how do we estimate this effort?
- *Reasonable idea:* number of binary ("yes"-"no") questions.
- *Resulting strategy:* binary search.
- *Idea:* we choose a question for which the answer is "yes" for exactly half of the remaining alternatives.
- *Property:* we need $\log_2(N)$ questions to select one of N alternatives.

2. Experts Are Usually More Comfortable with "Yes" Answers

- *In practice:* most people feel more comfortable answering "yes" than "no".
- *Fact:* the expert's time is valuable.
- *Consequence:* an expert is usually called after competent people tried to solve the problem.
- *Expected situation:* the expert mostly confirms their preliminary solutions.
- *Consequence:* most expert's answers are "yes".
- *Binary search case:* half of the answers are "no"s.
- *Meaning:* half of the previous decisions were wrong.
- *Expert's conclusion:* no competent people tried this problem – so his/her valuable time was wasted.

3. Experts Are Usually More Comfortable with "Yes" Answers (cont-d)

- *Situation:* a knowledge engineer interviews the expert.
- *First alternative:* most answers are "yes"; meaning:
  - the knowledge engineer already has some preliminary knowledge of the area, and
  - he/she is appropriately asking these questions to improve this knowledge.
- *Binary search:* half of the answers are "no" (same as for random questions); interpretation:
  - the knowledge engineer did not bother to get preliminary knowledge;
  - the highly skilled expert is inappropriately used to answer questions
  - which could be answered by consulting a textbook.

4. Problem

- *Reminder:* experts prefer "yes" answers.
- *Additional phenomenon:*
  - the larger the number of negative answers,
  - the more discomfort the expert will experience, and
  - the larger effort he will have to make to continue this interview.
- *Previous objective:* minimize the total number of questions.
- *More appropriate objective:* minimize the effort of an expert.
- *How to describe the effort:* assign more weight to "no" answers than to "yes" ones.
- *What we do:* find a search procedure which attains this objective.

5. How to Describe Different Search Procedures

- Let S be the set of N alternatives.
- We denote "yes" as 1, "no" as 0, so each sequence of answers ω is a binary sequence.
- To describe a search procedure, we must have:
  - the set Ω of possible answer sequences ω, and
  - a mapping A which maps each $\omega \in \Omega$ to the set $A(\omega)$ of all alternatives which are consistent with ω.
- *Formally:* $A(\Lambda) = S$, and for every $\omega \in \Omega$:
  - if $|A(\omega)| = 1$, then no extension of ω belongs to Ω;
  - otherwise, $\omega 0 \in \Omega$, $\omega 1 \in \Omega$, and we have

$$A(\omega) = A(\omega 0) \cup A(\omega 1),\ A(\omega 0) \cap A(\omega 1) = \emptyset,$$

$$A(\omega 0) \neq \emptyset, A(\omega 1) \neq \emptyset.$$

6. How to Gauge Different Search Procedures

- Let $P = (\Omega, A)$ be a search procedure.
- Let W_0 be the cost of "no" answer, and $W_1 < W_0$ be the cost of the "yes" answer.
- For $a \in \Omega$, let $\omega(a, P) = \omega_1\omega_2 \ldots \omega_k$ denote the sequence of answers which leads to a.
- The *cost* $W(\omega(a, P))$ of finding a is defined as

  $W(\omega(a,P)) = W(\omega_1\omega_2\ldots\omega_k) = W_{\omega_1} + W_{\omega_2} + \ldots + W_{\omega_k}$.
- The *effort* of a procedure is defined as the largest of its costs:
$$E(P) = \max_{a \in S} W(\omega(a, P)).$$
- *Objective:* find a procedure P_{opt} with the smallest possible effort:
$$E(P_{\text{opt}}) = T(N) \stackrel{\text{def}}{=} \min_{P} E(P).$$

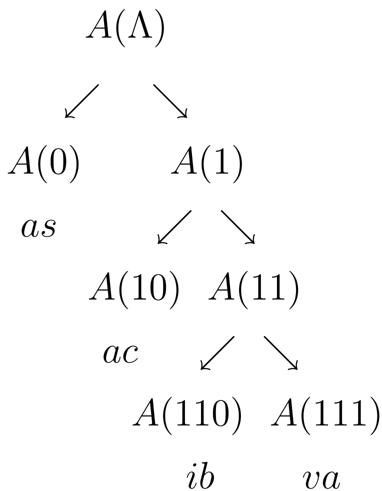
7. Example 1: Binary Search (Optimal for $W_0 = W_1$)

- *Situation:* a doctor chooses between $N = 4$ possible analgetics:
  - aspirin (as),
  - acetaminophen (ac),
  - ibuprofen (ib), and
  - valium (va).
- *Binary search:*

$$\begin{array}{ccccccc} & & & A(\Lambda) & & & \\ & & \swarrow & & \searrow & & \\ & A(0) & & & & A(1) & \\ \swarrow & & \searrow & & \swarrow & & \searrow \\ A(00) & & A(01) & & A(10) & & A(11) \\ as & & ac & & ib & & va \end{array}$$

8. Example 2: A Search Procedure Which Is Better Than Binary ($W_0 > W_1$)

- When $W_1 = 1$ and $W_0 = 3$, the effort of the binary search is 6.
- We can decrease the effort to 5 by applying the following alternative procedure:



9. Description of the Optimal Search Procedure

- *Auxiliary result:*

$$T(N) = \min_{0<N_+<N} \{\max\{W_1 + T(N_+), W_0 + T(N - N_+)\}\}.$$

- *Conclusion:* we can consequently compute $T(1)$, $T(2)$, $\ldots$, $T(N)$ in time $N \cdot O(N) = O(N^2)$.

- *Notation:* let $N_+(N)$ be the value where the minimum is attained.

- *Optimal procedure:* for each sequence ω with $n \stackrel{\text{def}}{=} |A(\omega)| > 1$:
  - we assign $N_+(n)$ values to the "yes" case $A(\omega 1)$;
  - we assign the remaining $n - N_+(n)$ values to the "no" case $A(\omega 0)$.

10. Example: $N = 4$, $W_0 = 3$, and $W_1 = 1$

- We take $T(1) = 0$. Then,

  $$T(2) = \min_{0<N_+<2} \{\max\{1 + T(N_+), 3 + T(2 - N_+)\}\} =$$

  $$\max\{1+T(1), 3+T(1)\} = \max\{1, 3\} = 3, \text{ with } N_+(2) = 1.$$

- $T(3) = 4$, with min attained for $N_+(3) = 2$.
- $T(4) = 5$, with min attained for $N_+(4) = 3$.
- *Optimal procedure:*
  - since $N_+(4) = 3$, we divide 4 elements $A(\Lambda)$ into a 3-element set $A(1)$ and a 1-element set $A(0)$;
  - since $N_+(3) = 2$, we divide 3 elements $A(1)$ into a 2-element set $A(11)$ and a 1-element set $A(10)$;
  - since $N_+(2) = 1$, we divide 2 elements $A(10)$ into a 1-element set $A(101)$ and a 1-element set $A(100)$.
- *Observation:* this is the procedure from Example 2.

11. Asymptotically Optimal Search Procedure

- *We described:* optimal search procedure:

$$E(P_{\text{opt}}) = T(N) \stackrel{\text{def}}{=} \min_P E(P).$$

- *Property:* P_{opt} takes time $\approx N^2$.
- *Problem:* for large N, time N^2 is too large.
- *Alternative:* asymptotically optimal procedure, with $E(P_a) \le T(N) + C$ for some constant $C > 0$.
- *Asymptotically optimal search procedure:*
  - find α such that $\alpha + \alpha^w = 1$, where $w \stackrel{\text{def}}{=} W_0/W_1$;
  - for each ω with $n \stackrel{\text{def}}{=} |A(\omega)| > 1$, assign $\lfloor \alpha \cdot n \rfloor$ values to the "yes" case $A(\omega 1)$;
  - assign the remaining values to the "no" case $A(\omega 0)$.

12. Examples

- *Reminder:* we find α such that $\alpha + \alpha^w = 1$, where $w \stackrel{\text{def}}{=} W_0/W_1$.
- *Example 1:*
  - *description:* $W_0 = W_1 = 1$, so $w = 1$;
  - *equation:* $\alpha + \alpha = 1$;
  - *solution:* $\alpha = 0.5$;
  - *resulting algorithm:* binary search.
- *Example 2:*
  - *description:* $W_0 = 2$, $W_1 = 1$, so $w = 2$;
  - *equation:* $\alpha + \alpha^2 = 1$;
  - *solution:* $\alpha = \dfrac{\sqrt{5} - 1}{2} \approx 0.618$ is the golden ratio;
  - *resulting algorithm:* asymmetric search.

13. Average Case vs. Worst Case

- *Reminder:* we gauged each procedure P by its worst-case effort $E(P)=\max\limits_{a\in S} W(\omega(a,P))$.
- *Alternative:* use the average-case effort

$$E^a(P) \stackrel{\text{def}}{=} \frac{1}{N}\cdot \sum_{a\in S} W(\omega(a,P)).$$

- *Problem:* find P_{opt} for which

$$E^a(P_{\text{opt}}) = T^a(N) \stackrel{\text{def}}{=} \min_P E^a(P).$$

- *Auxiliary result:* $T^a(N) = \min\limits_{0<N_+<N} \left\{\frac{N_+}{N}\cdot (W_1+T^a(N_+)) + \right.$

$$\left.\frac{N-N_+}{N}\cdot(W_0+T^a(N-N_+))\right\}.$$

- *Notation:* let $N^a_+(N)$ be the value where the minimum is attained.

14. Average Case: Optimal and Asymptotically Optimal Search Procedures

- *Optimal procedure:* for each ω with $n \stackrel{\rm def}{=} |A(\omega)| > 1$,
  - assign $N^A_+(n)$ values to the "yes" case $A(\omega 1)$;
  - assign the remaining values to the "no" case $A(\omega 0)$.
- *Asymptotically optimal procedure:*
  - find $K^a \ge 0$ and α^a for which:

$$\alpha^a \cdot W_1 + (1-\alpha^a) \cdot W_0 + K^a \cdot (\alpha^a \cdot \log_2(\alpha^a) + (1-\alpha^a) \cdot \log_2(1-\alpha^a)) = 0;$$

$$W_0 - W_1 = K^a \cdot (\log_2(\alpha^a) - \log_2(1-\alpha^a));$$

  - for each ω with $n \stackrel{\rm def}{=} |A(\omega)| > 1$, assign $\lfloor \alpha^a \cdot n \rfloor$ values to the "yes" case $A(\omega 1)$;
  - assign the remaining values to the "no" case $A(\omega 0)$.

15. Observation

- *Reminder:* one of the equations is

$$K^a = \frac{\alpha^a \cdot W_1 + (1 - \alpha^a) \cdot W_0}{-\alpha^a \cdot \log_2(\alpha^a) - (1 - \alpha^a) \cdot \log_2(1 - \alpha^a)}.$$

- *Reminder:* α^a and $1 - \alpha^a$ are the probabilities of the "yes" and "no" answers.

- *First conclusion:* the numerator $\alpha^a \cdot W_1 + (1 - \alpha^a) \cdot W_0$ is the average effort.

- *Second conclusion:* the denominator

$$-\alpha^a \cdot \log_2(\alpha^a) - (1 - \alpha^a) \cdot \log_2(1 - \alpha^a)$$

is the entropy of the probability distribution.

- *General conclusion:*

$$K^a = \frac{\text{average effort}}{\text{entropy of the probability distribution}}.$$

16. Acknowledgments

This work was supported in part by:

- by National Science Foundation grants HRD-0734825, EAR-0225670, and EIA-0080940,
- by Texas Department of Transportation grant No. 0-5453,
- by the Japan Advanced Institute of Science & Technology (JAIST) Int'l Joint Research Grant 2006-08, and
- by the Max Planck Institut für Mathematik.