How Intelligence Community Interprets Imprecise (Fuzzy) Words, and How to Justify This Empirical-Based Interpretation

Olga Kosheleva[1] and Vladik Kreinovich[2]

[1] Department of Teacher Education
University of Texas at El Paso, El Paso, TX 79968, USA
olgak@utep.edu
[2] Department of Computer Science
University of Texas at El Paso, El Paso, TX 79968, USA
vladik@utep.edu

Abstract. To provide a more precise meaning to imprecise (fuzzy) words like "probable" or "almost certain", researchers analyzed how often intelligence predictions hedged by each corresponding word turned out to be true. In this paper, we provide a theoretical explanation for the resulting empirical frequencies.

1 How Intelligence Community Interprets Imprecise (Fuzzy) Words

Need to interpret imprecise (fuzzy) words. A large portion of expert's knowledge is formulated by experts by using imprecise (fuzzy) words from natural language, such as "most probably", "small"', etc.

We humans understand such words, but computers have big trouble understanding such a knowledge. Computers are designed to process numbers, not words. It is therefore necessary to develop techniques that would translate such words into the language of numbers. This need was one of the main motivations behind Lotfi Zadeh's idea of fuzzy logic; see, e.g., [1, 5, 6, 8, 9, 14].

Why intelligence community needs to interpret imprecise (fuzzy) words. Based on different pieces of intelligence, intelligence analysts estimate the possibility of different scenarios. Their estimates usually come in terms of imprecise (fuzzy) words from natural language such as "almost certain", "probable", etc.

While the words themselves are imprecise, we can make their meaning more precise if we analyze all the situations in which the expert used the corresponding word – and count the frequency of cases in which the corresponding event actually happened. We expect that for the cases when the experts were almost certain, the corresponding frequency would be higher than for situations in which the experts simply stated that the corresponding event is probable.

Results of such analysis. This analysis was indeed undertaken at the US Central Intelligence Afgency (CIA), under the leadership of Sherman Kent; see, e.g., [3, 4] (see also [2, 12, 13]). This analysis has shown that the imprecise words describing expert's degree of certainty can be divided into seven groups. Within each group, words are practically synonymous to each other. The frequencies corresponding to a typical word from each group are as follows:

certain	100%
almost certain	93%
probable	75%
chances about even	50%
probably not	30%
almost certainly not	7%
impossible	0%

Here are example of synonyms:

- for "almost certain":
  - virtually certain,
  - all but certain,
  - high;y probable,
  - highly likely,
  - odds (or chances) overwhelming;
- for "possible":
  - conceivable,
  - could,
  - may,
  - might,
  - perhaps;
- for 50-50:
  - chances about even,
  - chances a little better (or less) than even;
  - improbable,
  - unlikely;
- for "probably not":
  - we believe that not,
  - we estimate that not,
  - we doubt,
  - doubtful;
- for "almost certainly not":
  - virtually impossible,
  - almost impossible,
  - some slight chance,
  - highly doubtful.

What is clear and what is not clear about this empirical result. The fact that we got exactly seven different categories in in perfect agreement with the well-known "seven plus minus two law" (see, e.g., [7, 10]) according to which human usually divide everything into seven plus minus two categories – with the average being exactly seven.

What is not clear is why namely the above specific probabilities are associated with seven terns, and not, e.g., more naturally sounding equidistant frequencies

$$0,\ \frac{1}{6},\ \frac{2}{6}\left(=\frac{1}{3}\right),\ \frac{3}{6}\left(=\frac{1}{2}\right),\ \frac{4}{6}\left(=\frac{2}{3}\right),\ \frac{5}{6},\ 1.$$

What we do in this paper. In this paper, we provide a theoretical explanation for the above empirical frequencies.

2 Towards a Theoretical Explanation for Empirical Frequencies

We make decisions based on finite number of observations. Crudely speaking, expert's estimates are based on his/her past experience. At any given moment of time, an expert has observed a finite number of observations. Let us denote this number by n.

If the actual probability of an event is p, then, for large n, the observed frequency is approximately normally distribution, with mean $\mu = p$ and standard deviation

$$\sigma = \sqrt{\frac{p \cdot (1-p)}{n}};$$

see, e.g., [11].

For two different processes, with probabilities p and p', the difference between the corresponding frequencies is also normally distributed, with mean $d \stackrel{\text{def}}{=} p - p'$ and standard deviation

$$\sigma_d = \sqrt{\sigma^2 + (\sigma')^2},$$

where σ is as above and

$$\sigma' = \sqrt{\frac{p' \cdot (1-p')}{n}}.$$

In general, for a normal distribution,all the values are:

- within the 2-sigma interval $[\mu - 2\sigma, \mu + 2\sigma]$ with probability $\approx 90\%$;
- within the 3-sigma interval $[\mu - 3\sigma, \mu + 3\sigma]$ with probability $\approx 99.9\%$;
- within the 6-sigma interval $[\mu - 6\sigma, \mu + 6\sigma]$ with probability $\approx 1 - 10^{-8}$, etc.

Whatever level of confidence we need, for appropriate k_0, all the value are within the interval $[\mu - k_0 \cdot \sigma, \mu + k_0 \cdot \sigma]$ with the desired degree of confidence.

Thus:

– if $|p - p'| \le k_0 \cdot \sigma_d$, then the zero difference between frequencies belongs to the k_0-sigma interval
$$[\mu - k_0 \cdot \sigma_d, \mu + k_0 \cdot \sigma_d]$$
and thus, it is possible that we will observe the same frequency in both cases;
– on the other hand, if $|p-p'| > k_0 \cdot \sigma_d$, this means that the zero difference between the frequencies is no longer within the corresponding k_0-sigma interval and thus, the observed frequencies are always different; so, by observing the corresponding frequencies, we can always distinguish the resulting probabilities.

Natural idea. Since we cannot distinguish close probabilities, we have a finite number of distinguishable probabilities. It is natural to try to identify them with the above empirically observed probabilities.

From the qualitative idea to precise formulas. For each value p, the smallest value $p' > p$ which can be distinguished from p based on n observations is the value $p' = p + \Delta p$, where $\Delta p = k_0 \cdot \sigma_d$. When $p \approx p'$, we have $\sigma \approx \sigma'$ and thus,
$$\sigma_m \approx \frac{\sqrt{2p \cdot (1-p)}}{n}.$$
So,
$$\Delta p = k_0 \cdot \frac{\sqrt{2p \cdot (1-p)}}{n}.$$

By moving all the terms connected to p to the left-hand side of this equality, we get the following equality:
$$\frac{\Delta p}{\sqrt{p \cdot (1-p)}} = k_0 \cdot \sqrt{\frac{2}{n}}. \quad (1)$$

By definition, the Δp is the difference between one level and the next one. Let us denote the overall number of levels by L. Then, we can associate:

– Level 0 with number 0,
– Level 1 with number $\frac{1}{L-1}$,
– Level 2 with number $\frac{2}{L-1}$, etc.,
– until we reach level $L - 1$ to which we associate the value 1.

Let $v(p)$ is the value corresponding to probability p. In these terms, for the two neighboring values, we get
$$\Delta v = \frac{1}{L-1},$$
thus $1 = L \cdot \Delta v$, and the formula (1) takes the form
$$\frac{\Delta p}{\sqrt{p \cdot (1-p)}} = k_0 \cdot \sqrt{\frac{2}{n}} \cdot (L-1) \cdot \Delta v,$$

i.e., the form

$$\frac{\Delta p}{\sqrt{p\cdot(1-p)}}=c\cdot\Delta v,$$

where we denoted

$$c\stackrel{\text{def}}{=}k_0\cdot\sqrt{\frac{2}{n}}\cdot(L-1).$$

The differences Δp and Δv are small, so we can approximate the above difference equation by the corresponding differential equation

$$\frac{dp}{\sqrt{p\cdot(1-p)}}=c\cdot dv.$$

Integrating both sides, we conclude that

$$\int\frac{dp}{\sqrt{p\cdot(1-p)}}=c\cdot v.$$

The integral in the left-hand side can be explicitly computed if we substitute $p=\sin^2(t)$ for some auxiliary quantity t. In this case, $dp=2\cdot\sin(t)\cdot\cos(t)\cdot dt$, and $1-p=1-\sin^2 t=\cos^2(t)$, thus

$$\sqrt{p\cdot(1-p)}=\sqrt{\sin^2(t)\cdot\cos^2(t)}=\sin(t)\cdot\cos(t).$$

Hence,

$$\frac{dp}{\sqrt{p\cdot(1-p)}}=\frac{2\sin(t)\cdot\cos(t)\cdot dt}{\sin(t)\cdot\cos(t)}=2dt,$$

so

$$\int\frac{dp}{\sqrt{p\cdot(1-p)}}=2t,$$

and the above formula takes the form

$$t=\frac{c}{2}\cdot v.$$

Thus,

$$p=\sin^2(t)=\sin^2\left(\frac{c}{2}\cdot v\right).$$

We know that the highest level of certainty $v=1$ corresponds to $p=1$, so

$$\sin^2\left(\frac{c}{2}\right)=1,$$

hence

$$\frac{c}{2}=\frac{\pi}{2}$$

and $c=\pi$.

Finally, we arrive at the following formula for the dependence on p on v:

$$p=\sin^2\left(\frac{\pi}{2}\cdot v\right).$$

In our case, we have 7 levels: Level 0, Level 1, …, until we reach Level 6. Thus, the corresponding values of v are $\frac{i}{6}$. So:

- for Level 0, we have $v = 0$, hence

$$p = \sin^2\left(\frac{\pi}{2}\cdot 0\right) = 0;$$

- for Level 1, we have $v = \frac{1}{6}$, so we have

$$p = \sin^2\left(\frac{\pi}{2}\cdot\frac{1}{6}\right) = \sin^2\left(\frac{\pi}{12}\right) \approx 6.7\% \approx 7\%;$$

- for Level 2, we have $v = \frac{2}{6} = \frac{1}{3}$, so we have

$$p = \sin^2\left(\frac{\pi}{2}\cdot\frac{1}{3}\right) = \sin^2\left(\frac{\pi}{6}\right) = \sin^2(30^\circ) = (0.5)^2 = 0.25;$$

- for Level 3, we have $v = \frac{3}{6} = \frac{1}{2}$, so we have

$$p = \sin^2\left(\frac{\pi}{2}\cdot\frac{1}{2}\right) = \sin^2\left(\frac{\pi}{4}\right) = \sin^2(45^\circ) = \left(\frac{\sqrt{2}}{2}\right)^2 = 0.5;$$

- for Level 4, we have $v = \frac{4}{6} = \frac{2}{3}$, so we have

$$p = \sin^2\left(\frac{\pi}{2}\cdot\frac{2}{3}\right) = \sin^2\left(\frac{\pi}{3}\right) = \sin^2(60^\circ) = \left(\frac{\sqrt{3}}{2}\right)^2 = \frac{3}{4} = 0.75;$$

- for level 5, we have $v = \frac{5}{6}$, so we have

$$p = \sin^2\left(\frac{\pi}{2}\cdot\frac{5}{6}\right) = \sin^2\left(\frac{5\pi}{12}\right) \approx 0.93;$$

item finally, for Level 6, we have we have $v = 1$, hence

$$p = \sin^2\left(\frac{\pi}{2}\cdot 1\right) = 1^2 = 1;$$

Conclusion. We have an *almost perfect* match.

The only difference is that, for Level 2, we get 25% instead of 30%. However, since the intelligence sample was not big, we can probably explain this difference as caused by the small size of the sample.

Acknowledgments

This work was supported in part by the US National Science Foundation grant HRD-1242122.

References

1. R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.
2. Central Intelligence Agency, Center for the Study of Intelligence, *Sherman Kent and the Profession of Intellogent Analysis*, Washington, DC, 2002.
3. S. Kent, “Rsimates and Influence”, *Studies in Intelloigence*, Summer 1968; https://www.cia.gov/library/center-for-the-study-of-intelligence/ csi-publications/books-and-monographs/ sherman-kent-and-the-board-of-national-estimates-collected-essays/ 4estimates.html
4. S. Kent, “Words of estimative probability”, in [12]; https://www.cia.gov/library/center-for-the-study-of-intelligence/ csi-publications/books-and-monographs/ sherman-kent-and-the-board-of-national-estimates-collected-essays/6words.html
5. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
6. J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
7. G. A. Miller, “The magical number seven, plus or minus two: Some limits on our capacity for processing information”, *Psychological Review*, 1956, Vol. 63, No. 2, pp. 81–97.
8. H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
9. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
10. S. K. Reed, *Cognition: Theories and application*, Wadsworth Cengage Learning, Belmont, California, 2010.
11. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.
12. D. P. Steury (ed.), *Sharman Kent and the Board of National Estimates*, Central Intelligence Agency, Center for the Study of Intelligence, Washington, DC, 1994.
13. P. E. Tetlock and D. Gardner, *Superforecasting: The Art and Science of Prediction*, Broadway Books, New York, 2015.
14. L. A. Zadeh, “Fuzzy sets”, *Information and Control*, 1965, Vol. 8, pp. 338–353.