How Intelligence Community Interprets Imprecise Evaluative Linguistic Expressions, 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 evaluative linguistic expressions like "probable" or "almost certain", researchers analyzed how often intelligence predictions hedged by each corresponding evaluative expression turned out to be true. In this paper, we provide a theoretical explanation for the resulting empirical frequencies.

Keywords: evaluative linguistic expressions, numerical estimates of linguistic expressions, intelligence community, empirical frequencies, theoretical explanation

1 How Intelligence Community Interprets Imprecise Evaluative Linguistics Expressions

Need to interpret imprecise evaluative linguistic expressions. A large portion of expert's knowledge is formulated by experts who use imprecise evaluative expressions from natural language, such as "most probably", "small", etc.

We humans understand such evaluative expressions, but computers have big trouble understanding such a knowledge. Computers are designed to process numbers, not linguistic expressions. It is therefore necessary to develop techniques that would translate such evaluative expressions 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, 7, 9, 12, 14, 21].

Zadeh's pioneering ideas inspired many techniques for assigning numerical values to different evaluative linguistic expressions; see, e.g., [13] and references therein.

Why intelligence community needs to interpret imprecise evaluative linguistic expressions in numerical terms. The ultimate objective of intelligence estimates is to make decisions. According to the decision theory analysis (see, e.g., [3, 4, 8, 11, 15]), a rational person should select an action that maximizes the expected value of a special *utility function* (a function that describes a person's preferences). To compute the expected value of the utility function, we need to know the probabilities of different possible events. Thus, to make a decision, we need to estimate the probabilities of different possible consequences of each action.

Based on different pieces of intelligence, intelligence analysts estimate the possibility of different scenarios. Their estimates usually come in terms of imprecise evaluative expressions from natural language such as "almost certain", "probable", etc. To use these estimates in decision making, it is therefore desirable to come up with a probability corresponding to each such evaluative expression.

How intelligence community interprets imprecise evaluative linguistic expressions: main idea. A natural way to assign a probability value to each evaluative linguistic expression is as follows:

- we consider all situations in which the experts' prediction used the corresponding evaluative linguistic expression, and
- we consider the frequency with which these predictions turned out to be true.

For example, if among 40 predictions in which the experts used the expression "probable", the prediction turned out to be true in 30 cases, the corresponding frequency is $30/40 = 75\%$.

Main idea: what we expect. It is reasonable to expect that:

- the more confident the experts are,
- the higher should be the frequencies with which these predictions turn out to be right.

For example, 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 future event is probable.

Possibility to go beyond the main idea. It is worth mentioning that in situations where a sample is too small to provide a meaningful estimation of the frequency, we can use an alternative approach for providing numerical estimates for evaluative linguistic expression as described in [16]. In this alternative approach:

- We ask several experts to estimate the degree of confidence (subjective probability) corresponding to each of these expressions.
- For each expression, we then take the average of degrees provided by different experts as the (subjective) probability corresponding to this evaluative linguistic expression.
- The standard deviation of these degrees can then be used as gauging the accuracy of the average-based probability estimate.

How intelligence community interprets imprecise evaluative linguistic expressions: the resulting interpretation. The analysis along the lines of the above-described main idea was indeed undertaken at the US Central Intelligence Agency (CIA), under the leadership of Sherman Kent; see, e.g., [5, 6] (see also [2, 19, 20]). This analysis has shown that the imprecise evaluative linguistic expressions describing expert's degree of certainty can be divided into seven groups. Within each group, evaluative expressions have approximately the same frequency. The frequencies corresponding to a typical evaluative expression from each group are described in Table 1.

Table 1. Empirical frequencies corresponding to different evaluative linguistic expressions

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

Here are the groups of evaluative linguistic expressions as described in [5, 6]:

– the group containing the expression "almost certain" also contained the following expressions:
  - virtually certain,
  - all but certain,
  - highly probable,
  - highly likely,
  - odds (or chances) overwhelming;
– the group containing the expression "possible" also contained the following expressions:
  - conceivable,
  - could,
  - may,
  - might,
  - perhaps;
– the group containing the expression "50-50" also contained the following expressions:
  - chances about even,
  - chances a little better (or less) than even;
  - improbable,
  - unlikely;

- the group containing the expression "probably not" also contained the following expressions:
  - we believe that not,
  - we estimate that not,
  - we doubt,
  - doubtful;
- the group containing the expression "almost certainly not" also contained the following expressions:
  - 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 is in perfect agreement with the well-known "seven plus minus two law" (see, e.g., [10,17]) 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 terms, and not, e.g., more naturally sounding equidistant frequencies

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

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 distributed, with mean $\mu = p$ and standard deviation

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

see, e.g., [18].

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,

$$\sigma_m \approx \sqrt{\frac{2p \cdot (1 - p)}{n}}.$$

So,

$$\Delta p = k_0 \cdot \sqrt{\frac{2p \cdot (1 - p)}{n}}.$$

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 $\dfrac{1}{L-1}$,
– Level 2 with number $\dfrac{2}{L-1}$,
– ...
– 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 $1 = (L-1)\cdot \Delta v$, 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 $v = \dfrac{5}{6}$, so we have

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

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

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

Discussion. 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.

3 Conclusions

To gauge to what extend different future events are possible, experts often use evaluative linguistic expressions such as "probable", "almost certain", etc. Some predictions turn out to be true, some don't. A natural way to gauge the degree of confidence as described by a given evaluative expression is to analyze, out of all the prediction that used this expression, how many of them turned out to be true. Such an analysis was indeed performed by the intelligence community, and the corresponding empirical frequencies have been used to make expert's predictions more precise.

In this paper, we provide a theoretical explanation for the resulting empirical frequencies. This explanation is based on a natural probabilistic analysis of the corresponding situation.

Acknowledgments

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

The authors are greatly thankful to the anonymous referees for their valuable suggestions and to Dan Tavrov for his help.

References

1. Belohlavek R., Dauben J. W., Klir G. J.: 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 intelligent analysis, Washington, DC (2002).
3. Fishburn P. C.: Utility theory for decision making, John Wiley & Sons Inc., New York (1969).

4. Fishburn P. C.: Nonlinear preference and utility theory, The John Hopkins Press, Baltimore, Maryland (1988).
5. Kent S.: Esimates and Influence, Studies in Intelligence, Central Intelligence Agency, Washington, DC (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
6. Kent S.: Words of estimative probability, in [19];
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
7. Klir G., Yuan B.: Fuzzy sets and fuzzy logic, Prentice Hall, Upper Saddle River, New Jersey (1995).
8. Luce R. D., Raiffa R.: Games and decisions: introduction and critical survey, Dover, New York (1989).
9. Mendel J. M.: Uncertain rule-based fuzzy systems: introduction and new directions, Springer, Cham, Switzerland (2017).
10. Miller G. A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information, Psychological Review 63(2), 81–97 (1956).
11. Nguyen H. T., Kosheleva O., Kreinovich V.: Decision making beyond Arrow's "impossibility theorem", with the analysis of effects of collusion and mutual attraction, International Journal of Intelligent Systems 24(1), 27–47 (2009).
12. Nguyen H. T., Walker E. A.: A first course in fuzzy logic, Chapman and Hall/CRC, Boca Raton, Florida (2006).
13. Novák V.: Evaluative linguistic expressions vs. fuzzy categories?, Fuzzy Sets and Systems, 281, 81–87 (2015).
14. Novák, V., Perfilieva, I., Močkoř J.: Mathematical principles of fuzzy logic, Kluwer, Boston, Dordrecht (1999).
15. Raiffa H.: Decision Analysis, Addison-Wesley, Reading, Massachusetts (1970).
16. Rajati M. R., Mendel J. M.: Modeling linguistic probabilities and linguistic quantifiers using interval type-2 fuzzy sets, Proceedings of the 2013 Joint World Congress of the International Fuzzy Systems Association and the Annual Conference of the North American Fuzzy Information Processing Society IFSA/NAFIPS'2013, Edmonton, Canada, June 24–28, 2013, pp. 327–332 (2013).
17. Reed S. K.: Cognition: theories and application, Wadsworth Cengage Learning, Belmont, California (2010).
18. Sheskin D. J.: Handbook of parametric and nonparametric statistical procedures, Chapman and Hall/CRC, Boca Raton, Florida (2011).
19. Steury D. P. (ed.): Sharman Kent and the Board of National Estimates, Central Intelligence Agency, Center for the Study of Intelligence, Washington, DC (1994).
20. Tetlock P. E., Gardner, D.: Superforecasting: the art and science of prediction, Broadway Books, New York (2015).
21. Zadeh L. A.: Fuzzy sets, Information and Control, 8, 338–353 (1965).