How Intelligence Community Interprets Imprecise Evaluative Linguistic Expressions, and How to Justify This Empirical-Based Interpretation

Olga Kosheleva\textsuperscript{1} and Vladik Kreinovich\textsuperscript{2}

Departments of \textsuperscript{1}Teacher Education and \textsuperscript{2}Computer Science University of Texas at El Paso, El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu
1. Need to Interpret Imprecise Evaluative Linguistic Expressions

- Experts often use imprecise evaluative expressions from natural language, such as “most probably”, “small”.
- Computers have big trouble understanding such a knowledge.
- Computers are designed to process numbers, not linguistic expressions.
- It is therefore necessary to translate such evaluative expressions into numbers.
- This need was one of the main motivations behind Lotfi Zadeh’s idea of fuzzy logic.
- Zadeh’s pioneering ideas inspired many techniques for assigning numerical values to different expressions.
2. **Intelligence Community Needs to Interpret Evaluative Expressions in Numerical Terms**

- The ultimate objective of intelligence estimates is to make decisions.

- According to the decision theory analysis, a person’s preferences are described by a utility function.

- A rational person should select an action that maximizes the expected utility.

- To compute the expected value of the utility function, we need to know the probabilities of different events.

- Thus, to make a decision, we need to estimate the probabilities of different consequences of each action.

- Based on different pieces of intelligence, intelligence analysts estimate the possibility of different scenarios.
3. Intelligence Community Needs to Interpret Evaluative Expressions (cont-d)

- 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.
4. How Intelligence Community Interprets Imprecise Evaluative 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 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 \( \frac{30}{40} = 75\% \).
5. 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.
6. 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 linguistic expressions.

• In this alternative approach, for each expression:
  – We ask several experts to estimate the related degree of confidence (subjective probability).
  – We then take the average as the (subjective) probability corresponding to this expression.

• The standard deviation can then be used as gauging the accuracy of this estimate.
7. How Intelligence Community Interprets Imprecise Evaluative Linguistic Expressions

- Sherman Kent implemented the above idea at CIA.
- His analysis showed that expressions can be divided into 7 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 the table.

<table>
<thead>
<tr>
<th>Evaluative Expression</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>certain</td>
<td>100%</td>
</tr>
<tr>
<td>almost certain</td>
<td>93%</td>
</tr>
<tr>
<td>probable</td>
<td>75%</td>
</tr>
<tr>
<td>chances about even</td>
<td>50%</td>
</tr>
<tr>
<td>probably not</td>
<td>30%</td>
</tr>
<tr>
<td>almost certainly not</td>
<td>7%</td>
</tr>
<tr>
<td>impossible</td>
<td>0%</td>
</tr>
</tbody>
</table>
8. Resulting Groups of Evaluative Expressions

- 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 “possible” also contains:
  - conceivable,
  - could,
  - may,
  - might,
  - perhaps.
9. Groups of Evaluative Expressions (cont-d)

- The group containing “50-50” also contains:
  - chances about even,
  - chances a little better (or less) than even;
  - improbable,
  - unlikely.

- The group containing “probably not” also contains:
  - we believe that not,
  - we estimate that not,
  - we doubt,
  - doubtful.
10. Groups of Evaluative Expressions (cont-d)

- The group containing “almost certainly not” also contains:
  - virtually impossible,
  - almost impossible,
  - some slight chance,
  - highly doubtful.
11. What Is Clear And What Is Not Clear About This Empirical Result

- 7 categories is in agreement with $7 \pm 2$ law.
- According to this law, humans usually divide everything into $7 \pm 2$ categories – on average, 7.
- What is not clear is why namely the above specific probabilities are associated with seven terms.
- Why not equidistant frequencies

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

- In this talk, we provide a theoretical explanation for the above empirical frequencies.
12. 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 st. dev.
    \[ \sigma = \sqrt{\frac{p \cdot (1 - p)}{n}}. \]
- The difference between frequencies corr. to $p \neq p'$ is also normally distributed.
13. We Make Decisions Based on Finite Number of Observations (cont-d)

- Its mean is \( d \triangleq p - p' \) and standard deviation \( \sigma_d = \sqrt{\sigma^2 + (\sigma')^2} \), where \( \sigma' = \sqrt{\frac{p' \cdot (1 - p')}{n}} \).

- 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 \( [\mu - k_0 \cdot \sigma, \mu + k_0 \cdot \sigma] \)
  - with the desired degree of confidence.
14. We Make Decisions Based on Finite Number of Observations (cont-d)

- If \(|p - p'| \leq k_0 \cdot \sigma_d\), then the zero difference between frequencies belongs to the \(k_0\)-sigma interval
  \[ [\mu - k_0 \cdot \sigma_d, \mu + k_0 \cdot \sigma_d]. \]

- 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 0 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.
15. Natural Idea

• We cannot distinguish close probabilities.

• Thus, we have a finite number of distinguishable probabilities.

• It is natural to try to identify them with the above empirically observed probabilities.
16. 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
  \[ p' = p + \Delta p, \text{ where } \Delta p = k_0 \cdot \sigma_d. \]

- When $p \approx p'$, then $\sigma \approx \sigma'$ and $\sigma_m \approx \sqrt{\frac{2p \cdot (1 - p)}{n}}$.

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

- By definition, the $\Delta p$ is the difference between one level and the next one.
17. From Idea to Formulas (cont-d)

- 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 $\frac{2}{L-1}$,
  - $\ldots$
  - 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
  $\Delta v = \frac{1}{L-1}$. 
18. From Idea to Formulas (cont-d)

• Thus $1 = (L - 1) \cdot \Delta v$, and the above formula 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.:

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

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

• The differences $\Delta p$ and $\Delta v$ are small.

• So, we can approximate the above difference equation by a differential equation

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

• Integrating, we get

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

• This integral can be explicitly computed if we substitute $p = \sin^2(t)$ for an auxiliary $t$. 

From the Qualitative...
19. From Idea to Formulas (cont-d)

• 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

\[
\int \frac{dp}{\sqrt{p \cdot (1 - p)}} = 2t.
\]

• So, 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 \( p = 1 \).
20. From Idea to Formulas (cont-d)

- So $\sin^2 \left( \frac{\pi}{2} \right) = 1$, 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 $\frac{i}{6}$.

- 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

$$p = \sin^2 \left( \frac{\pi}{2} \cdot \frac{1}{6} \right) = \sin^2 \left( \frac{\pi}{12} \right) \approx 6.7\% \approx 7\%.$$
21. From Idea to Formulas (cont-d)

- 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

$$p = \sin^2 \left( \frac{\pi}{2} \cdot \frac{2}{3} \right) = \sin^2 \left( \frac{\pi}{3} \right) = \sin^2(60^\circ) = \left( \frac{\sqrt{3}}{2} \right)^2 = \frac{3}{4} = 0.75.$$
22. From Idea to Formulas (cont-d)

- For Level 5, we have \( v = \frac{5}{6} \), so we have

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

- Finally, for Level 6, we have \( v = 1 \), hence

\[
p = \sin^2 \left( \frac{\pi}{2} \cdot 1 \right) = 1^2 = 1.
\]
23. 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.
24. 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.
- The corresponding empirical frequencies have been used to make expert’s predictions more precise.
25. Conclusions (cont-d)

• In this talk, we provide a theoretical explanation for the resulting empirical frequencies.

• This explanation is based on a natural probabilistic analysis of the corresponding situation.
26. Acknowledgments

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