Quantifying Aleatoric and Epistemic Uncertainty in Machine Learning

Are Conditional Entropy and Mutual Information Appropriate Measures?

Lisa Wimmer\textsuperscript{1,3}, Yusuf Sale\textsuperscript{2,3}, Paul Hofman\textsuperscript{2,3}, Bernd Bischl\textsuperscript{1,3}, Eyke Hüllermeier\textsuperscript{2,3}

\textsuperscript{1}Department of Statistics, LMU Munich
\textsuperscript{2}Institute of Informatics, LMU Munich
\textsuperscript{3}Munich Center for Machine Learning (MCML)
Predictive uncertainty

Common understanding

Uncertainty arising from

\[ \vdash \text{imperfect data} \]
\[ \vdash \text{limited knowledge} \]

Implicit assumption

\[ f^* \]

\[ f : X \rightarrow Y \]

Uncertainty components

Total uncertainty \( \equiv \) aleatoric uncertainty + epistemic uncertainty

\[ \text{TU} \]
\[ \text{AU} \]
\[ \text{EU} \]

---

\(^1\)Hüllermeier and Waegeman (2021), Kendall and Gal (2017)
3+ levels

- **Level 0**: \( y \in \mathcal{Y} \) // no uncertainty
- **Level 1**: \( \theta \in \mathbb{P}(\mathcal{Y}) \) // uncertainty about \( y \mid x \sim AU \)
- **Level 2**: \( Q \in \mathbb{P}(\mathbb{P}(\mathcal{Y})) \) // uncertainty about \( \theta \sim EU \)
Intuition

Uncertainty components

\[ \text{Total uncertainty} = \text{aleatoric uncertainty} + \text{epistemic uncertainty} \]

\[ \text{TU} = \text{AU} + \text{EU} \]

\( \therefore \text{TU, EU maximal for total ignorance} \)

\( \therefore \text{EU} \rightarrow 0 \text{ for } n \rightarrow \infty \)

\( \therefore \text{AU} \equiv c \in \mathbb{R} \)
Entropy-based measures

TU – Shannon entropy\(^2\)

\[
H(Y) = H(\mathbb{E}_{Q}[Y|\theta]) = -\sum_y p(y) \cdot \log p(y)
\]

AU – conditional entropy

\[
H(Y|\Theta) = \mathbb{E}_{Q}[H(Y|\theta)] = \mathbb{E}_{Q}[-\sum_y p(y|\theta) \log p(y|\theta)]
\]

EU – mutual information

\[
I(Y, \Theta) = H(Y) - H(Y|\Theta) \quad // \text{uncertainty reduction in } Y
\]

\(^2\)Shannon (1948), Houlsby et al. (2011), Cover and Thomas (2006)
Fundamental relationship

\[ \text{Shannon entropy} = \text{conditional entropy} + \text{mutual information} \]

\[ TU = AU + EU \]
Fundamental relationship

Shannon entropy \( TU \) = conditional entropy \( AU \) + mutual information \( EU \)

Proposition 5. If \( EU \) and \( TU \) attain their respective maxima at the beginning of learning, and they are constructed to be on the same scale, then \( TU \) cannot decompose additively into \( EU \) and \( AU \) if \( AU \) is positive.
Desired formal properties

A0  TU, AU and EU are non-negative.

A1  EU vanishes for Dirac measures $Q = \delta_\theta$.

A2  EU and TU are maximal for $Q$ being the uniform distribution.

A3  If $Q'$ is a mean-preserving spread of $Q$, then $EU(Q') \geq EU(Q)$ (weak version) or $EU(Q') > EU(Q)$ (strict version); the same holds for TU.

A4  If $Q'$ is a center-shift of $Q$, then $AU(Q') \geq AU(Q)$ (weak version) or $AU(Q') > AU(Q)$ (strict version); the same holds for TU.

A5  If $Q'$ is a spread-preserving location shift of $Q$, then $EU(Q') = EU(Q)$. 
A paradoxical example

A2 EU and TU are maximal for Q being the uniform distribution.
A paradoxical example

A5 If $Q'$ is a spread-preserving location shift of $Q$, then $\text{EU}(Q') = \text{EU}(Q)$. 

<table>
<thead>
<tr>
<th>TU: 1.00, AU: 0.72, EU: 0.28</th>
<th>TU: 1.00, AU: 0.00, EU: 1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph 1" /></td>
<td><img src="image2.png" alt="Graph 2" /></td>
</tr>
<tr>
<td>TU: 1.00, AU: 0.96, EU: 0.04</td>
<td>TU: 0.97, AU: 0.89, EU: 0.08</td>
</tr>
<tr>
<td><img src="image3.png" alt="Graph 3" /></td>
<td><img src="image4.png" alt="Graph 4" /></td>
</tr>
</tbody>
</table>
Empirical evidence

CIFAR10, deep ensemble

uncertainty

sample size in % of total

ACC / ECE

ACC
ECE
AU
EU
TU
Main criticism

- Inadequacy of standard uncertainty measures
  - EU: counter-intuitive behavior; measure of conflict rather than ignorance
  - AU: estimation under intrinsic uncertainty from level 2
  - TU: loss of information due to marginalization

- Additivity: not possible for finite $n$ under A0–A5
A way forward

Better measures

- Axiomatic foundation
- Inter-level uncertainty propagation

Other representational frameworks

- Beyond classical probability theory

\(^3\)Hüllermeier et al. (2022), Sale et al. (2023), Dubois et al. (1996)
Stop by

Poster #374

Questions?
References

Figure on last page. https://www.pexels.com/photo/white-and-black-mountain-wallpaper-933054/, 2023.


Figure on title page. https://losslandscape.com/, 2023.


