Supplementary materials

Reproducibility

This section deals with the experiments presented in this document, with the different parameters for each experiment. Note that every datasets as well as the code in Python allowing to reproduce these experiments are available at: https://anonymous.4open.science/r/evidential-uncertainty-sampling-D453.

Uncertainty sampling

The three datasets are generated according to the following separation line:

1. $y = -3x$
2. $y = \sin(0.8x)$
3. $(x-0)^2 + (y-0)^2 = 6$

The uncertainties are computed on a grid of $60 * 50$ points (it is the same for every experiments). The used model is K-NN (K-Nearest Neighbors), with a probabilistic output and on the distance-weighted version available with scikit-learn [5] (every other parameters are the scikit-learn default parameters). The model is trained on the dataset and then try to predict the class of every data point of the grid by using 10 nearest neighbors. The uncertainty used is the least confidence measure given in equation (5) and the same parameters are used on Fisher's Iris dataset.

Example on real world dataset: An example on real data is presented in Figure 1, Fisher's Iris dataset is presented on two variables. The three Iris classes are shown in 1a and the areas of model uncertainty are shown in 1b, these are the regions where the model would be the most uncertain to predict the class of a new observation. This is summarized by the fact that if a new observation falls into a red zone, the model will have more difficulty predicting its class.

Epistemic and aleatoric uncertainties

The dataset is generated according to the separation line of equation $y = 0$ and the data points are generated, with more points on the left according to a binary logarithm $\log_2$.

The uncertainties are computed according to the same version of the K-Nearest Neighbors algorithm. The experimental process follows the one described in [4]. The formulas are deduced from equation (6) as follows:

$$\begin{aligned} \pi(1|x) &= \sup_{\theta \in \Theta} \min[\pi_\Theta(\theta), p_\theta(1|x) - p_\theta(0|x)], \\ \pi(0|x) &= \sup_{\theta \in \Theta} \min[\pi_\Theta(\theta), p_\theta(0|x) - p_\theta(1|x)]. \end{aligned} \tag{1}$$

That can be rewritten with $f(a) = 2a - 1$:

$$\begin{aligned}\pi(1|x) &= \sup_{\theta\in\Theta} \min[\pi_\Theta(\theta), f(p_\theta(1|x))],\\ \pi(0|x) &= \sup_{\theta\in\Theta} \min[\pi_\Theta(\theta), f(1-p_\theta(1|x))],\end{aligned} \tag{2}$$

with:

$$\pi_\Theta(\theta) = \frac{L(\theta)}{L(\hat{\theta})}, \tag{3}$$

For a Parzen window, as described in [4] and with p and n the number of positive and negative instances within the window, $L(\theta)$ is:

$$L(\theta) = \binom{p+n}{p}\theta^p(1-\theta)^n, \text{ and } \hat{\theta} = \frac{p}{p+n} \tag{4}$$

As we are not using a Parzen window but the distance-weighted K-NN, we take p the sum of the inverse distance for positive elements among the K neighbors, and n reciprocally the same for the negative class. Yet we have:

$$\begin{aligned}\pi(1|x) &= \sup_{\theta\in[0,1]} \min\left(\frac{\theta^p(1-\theta)^n}{\left(\frac{p}{p+n}\right)^p\left(\frac{n}{p+n}\right)^n}, 2\theta - 1\right),\\ \pi(0|x) &= \sup_{\theta\in[0,1]} \min\left(\frac{\theta^p(1-\theta)^n}{\left(\frac{p}{p+n}\right)^p\left(\frac{n}{p+n}\right)^n}, 1 - 2\theta\right).\end{aligned} \tag{5}$$

Brent's method is applied to find a local minimum (here the maximum) in the interval $\theta \in [0, 1]$. Note that in this paper, the experimental results are not there to compare a performance gap, but to support a new approach to sampling by uncertainty. And, in our case, we have deduced the following formula more appropriate, so it is this one that we use in the presented experiments:

$$\begin{aligned}\pi(1|x) &= \sup_{\theta\in[0,1]} \min\left(\frac{\theta^p(1-\theta)^n}{\left(\frac{p}{p+n}\right)^p\left(\frac{n}{p+n}\right)^n}, 2\theta - 1\right),\\ \pi(0|x) &= 1 - \sup_{\theta\in[0,1]} \min\left(\frac{\theta^p(1-\theta)^n}{\left(\frac{p}{p+n}\right)^p\left(\frac{n}{p+n}\right)^n}, 1 - 2\theta\right),\end{aligned} \tag{6}$$

and for the uncertainties:

$$\begin{aligned}\mathcal{U}_e(x) &= \min[\pi(1|x), \pi(0|x)] - 0.5,\\ \mathcal{U}_a(x) &= 1 - \max[\pi(1|x), \pi(0|x)].\end{aligned} \tag{7}$$

The computational complexity, the dependence on observations, the optimization phase, the use of likelihood and maximum likelihood make the model heavier. In the next two experiments and thus for the proposed models, a simple output-dependent calculation replaces all these steps.

Discord and non-specificity

The dataset is generated according to the separation line of equation $y = 0$ and the data points are added, with no observations in the subspace verifying $(x - 2.8)^2 + (y - 0)^2 < 5$ and with imprecise (or more ignorant) observations in the subspace verifying $(x - 4.5)^2 + (y - 0)^2 < 5$.

The evidential uncertainties are computed according to the Evidential K-Nearest Neighbors algorithm [1]. The parameters are those used in the γ-EK-NN version presented in [2], with $K = 10$ nearest neighbors, $\beta = 2$, $\alpha = 1$ and the formula (12) is used with $\lambda = 0.5$, as suggested in [3]. When λ varies the values taken are $\lambda = 0.1$ and $\lambda = 0.9$. A difference is made with respect to the Evidential K-Nearest Neighbors model, instead of the conjunctive combination of Demster, a combination by average of the masses (3) is used (which allows to pass α to 1, instead of $\alpha < 1$).

Evidential epistemic uncertainty

The dataset is generated according to three gaussian scikit-learn blobs [5] with imprecise (or more ignorant) observations in the subspace verifying $(2.3 - x)^2 + (3 - y)^2 < 2$.

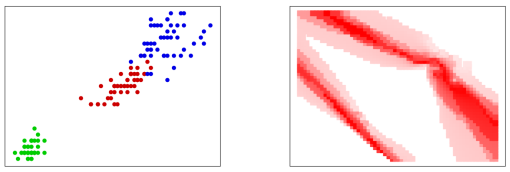
The evidential epistemic uncertainty is computed according to the same Evidential K-Nearest Neighbors with $K = 6$ nearest neighbors, $\beta = 2$, $\alpha = 1$ and the formula (19) is used to compute both the proposed evidential epistemic and aleatoric uncertainties.

Non-regression for advanced methods

This section shows that for the models presented, whether epistemic uncertainty or the proposition to use Klir uncertainty and evidential epistemic uncertainty, none is reductive regarding the (4) and (5) classical measures. Effectively it offers similar areas of uncertainty as the one used is section 2 when epistemic, aleatoric and evidential uncertainties are indistinguishable. Figures 2, 3 and 4 show the areas of uncertainty on the datasets introduced in Figure 1. The epistemic uncertainty, Klir uncertainty and the evidential epistemic uncertainty offers, roughly, the same areas of uncertainty as the one used in Figure 1.

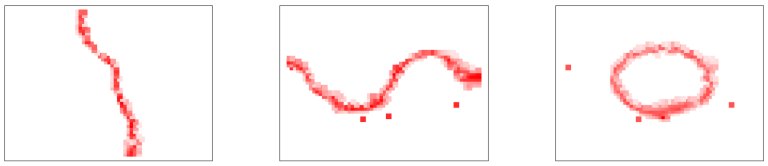
References

1. Denœux, T.: A k-nearest neighbor classification rule based on dempster-shafer theory. Systems, Man and Cybernetics, IEEE Transactions on **219** (1995)
2. Hoarau, A., Martin, A., Dubois, J.C., Le Gall, Y.: Imperfect labels with belief functions for active learning. In: Belief Functions: Theory and Applications. Springer International Publishing (2022)
3. Klir, G.J., Wierman, M.J.: Uncertainty-based information: Elements of generalized information theory. In: Springer-Verlag (1998)
4. Nguyen, V.L., Shaker, M.H., Hüllermeier, E.: How to measure uncertainty in uncertainty sampling for active learning. Mach. Learn. **111**, 89–122 (jan 2022)
5. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)



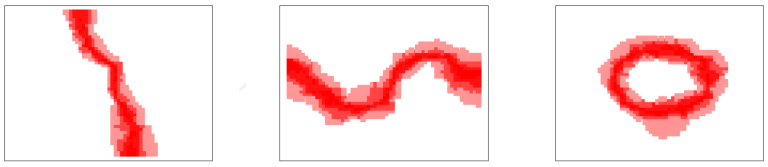
(a) Iris Dataset (b) Uncertainty

Fig. 1: Fisher's Iris dataset on two descriptive variables, petal length and width, with model uncertainties. Setosa in green, Versicolor in red and Virginica in blue.



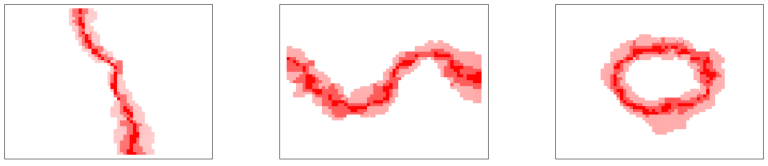
(a) (b) (c)

Fig. 2: From left to right, the areas of uncertainty corresponding to the datasets 1a, 1b and 1c according to the epistemic uncertainty.



(a) (b) (c)

Fig. 3: From left to right, the areas of uncertainty corresponding to the datasets 1a, 1b and 1c according to Klir uncertainty resulting from the discord and the non-specificity.



(a) (b) (c)

Fig. 4: From left to right, the areas of uncertainty corresponding to the datasets 1a, 1b and 1c according to the evidential epistemic uncertainty.