University Defence Research Collaboration (UDRC)
Signal Processing in a Networked Battlespace

UNIVERSITY OF SURREY

CVSSP Centre for Vision, Speech & Signal Processing

Incongruence Detection for Statistical Anomaly Detection

Josef Kittler, Cemre Zor and Wenwu Wang
Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, Guildford, GU2 7XH
{j.kittler, c.zor, w.wang}@surrey.ac.uk

Incongruence Detection

Aims to aid the detection of anomaly in sensor data processed by a complex decision making system. Focuses on:

- Comparing the outputs of two classifiers with a view to detecting statistical anomaly in sensor data
  - The nature/nuance of anomaly should subsequently be identified based on a detailed analysis of the classifier outputs
- Analysing measures of classifier incongruence, Histogram Consistency and Similarity Tests, Bayesian Surprise
- Development of alternative methods which focus on the dominant hypotheses flagged by two experts: Delta-Max (Δ_{max}) and Delta-Avg (Δ_{avg})

Incongruence Detection Background

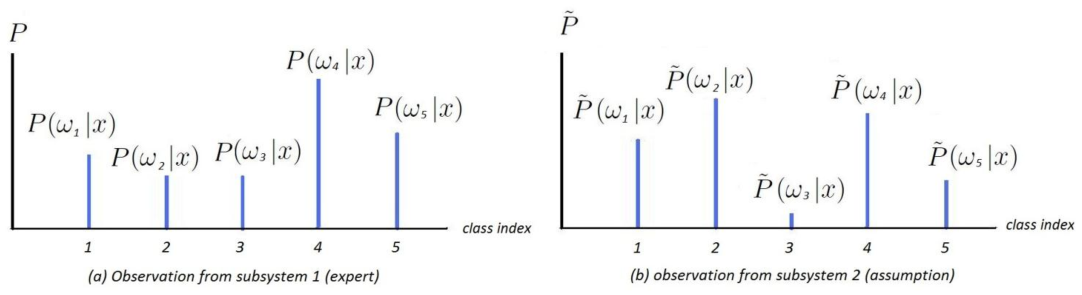
- Required ingredients
  - Incongruence measure
  - Estimate incongruence measure distribution
  - Statistical hypothesis testing threshold
- Existing incongruence measures
  - Histogram Consistency/Similarity Tests
    - Testing if two histograms are drawn from the same distribution, using shape analysis and statistics
    - E.g. Chi-square, Cramer-von-Mises, Kolmogorov Smirnov Tests
    - No single "best" test for all application
  - Bayesian Surprise (BS)
    - The Kullback-Leibler divergence between two expert distributions

$$\Delta_{BS} = \sum_{j=1}^{r} \tilde{P}(\omega_j|x) \log \frac{\tilde{P}(\omega_j|x)}{P(\omega_j|x)}$$

  - Cons: Divergence to infinity and non-symmetrical behaviours

Assumption: Different subsystems voice independent opinions about the strengths of various hypothesis

- Incongruence is to be detected for each of the outputs of different subsystems



Δ_{max} and Δ_{max}^{dm} Measures

Alternative methods which focus on the dominant hypotheses flagged by the two experts

$$\Delta_{max} = \frac{1}{2}[|\tilde{P}(\tilde{\mu}|x) - P(\tilde{\mu}|x)| + |\tilde{P}(\mu|x) - P(\mu|x)|]$$

where $\mu = \arg max_w P(w|x)$ and $\tilde{\mu} = \arg max_w \tilde{P}(w|x)$.

- Symmetric, eliminates the clutter injected by the non-dominant classes, does not diverge to infinity but confined to the interval (0,1).
- One undesirable property: When the classifiers vote for the same class, second term is identical to first term.
  - This scenario doubles the surprise measure, and masks the scenario when the favoured hypotheses of the two experts differ.
- Solution: An update on Δ_{max}: Δ_{max}^{dm}

$$\Delta_{max}^{dm} = \frac{1}{2}\max\{[|\tilde{P}(\mu|x) - P(\mu|x)| + \delta(\mu,\tilde{\mu})|\tilde{P}(\tilde{\mu}|x) - \tilde{P}(\mu|x)|],$$
$$[|P(\tilde{\mu}|x) - \tilde{P}(\tilde{\mu}|x)| + \delta(\mu,\tilde{\mu})|P(\mu|x) - P(\tilde{\mu}|x)|]\}$$

where $1 - \delta(\mu,\tilde{\mu}) = \begin{cases} 1 & if \;\; \mu = \tilde{\mu} \\ 0 & if \;\; \mu \neq \tilde{\mu} \end{cases}$

- Δ_{max}^{dm} is magnified if the two classifiers support distinct dominant hypothesis.

Δ_{avg} Measure

$$\begin{aligned}\Delta_{avg} \;\; = \tfrac{1}{2}\{&|P(\mu|x) - \tilde{P}(\mu|x)| + \delta(\mu,\tilde{\mu})|\tilde{P}(\tilde{\mu}|x) - \tilde{P}(\mu|x)| \\ &+|\tilde{P}(\tilde{\mu}|x) - P(\tilde{\mu}|x)| + \delta(\mu,\tilde{\mu})|P(\mu|x) - P(\tilde{\mu}|x)|\}\end{aligned}$$

- Similar properties with those of Δ_{max}.

Error Sensitivity Analysis

- A posteriori probabilities estimated by the two classifiers are subject to estimation errors ($P(w|x) + \eta_w(x)$).
  - Assumption: Errors are normally distributed with zero mean and σ stdev.
  - However, the following conditions must also be satisfied:

$$\sum_{i}^{m} \eta_\omega(\mathbf{x}) = 0 \quad \text{and} \quad 0 \leq \eta_\omega(\mathbf{x}) + P(\omega|\mathbf{x}) \leq 1$$

  - We adopt a new error distribution, p', which is a clipped normal distribution.

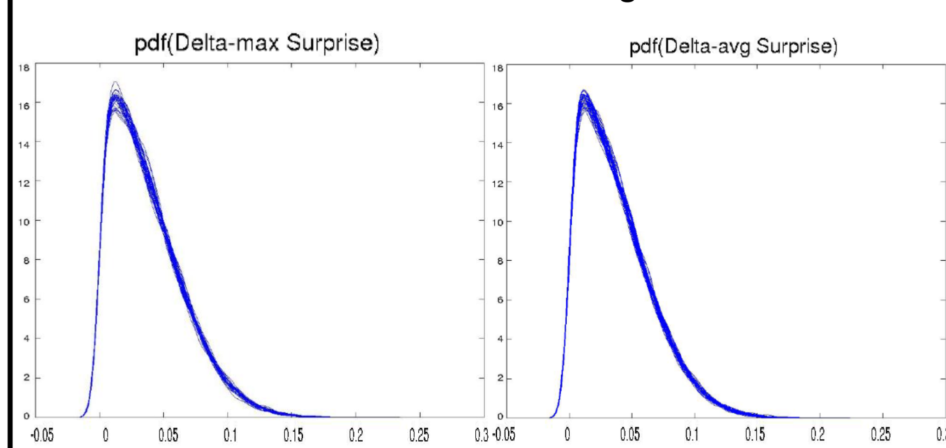
$$P \leq \;\; 0.5$$
$$p'(\eta) = \begin{cases} 0 & \eta < -P \\ p(\eta) + p(\,2P - \eta) & \eta \geq -P \end{cases}$$
$$P > \;\; 0.5$$
$$p'(\eta) = \begin{cases} 0 & \eta > 1 - P \\ p(\eta) + p(2 - 2P - \eta) & \eta \leq 1 - P \end{cases}$$

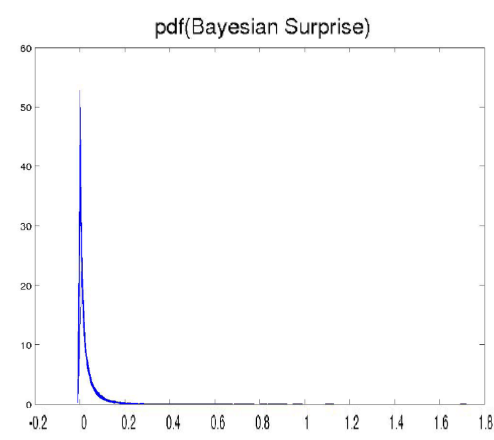
Experiments

Surprise distribution with error sensitivity analysis for different scenarios, using 2 and 3 class problems, is evaluated. BS, Δ_{max}^{dm} and Δ_{avg} measures are utilized.

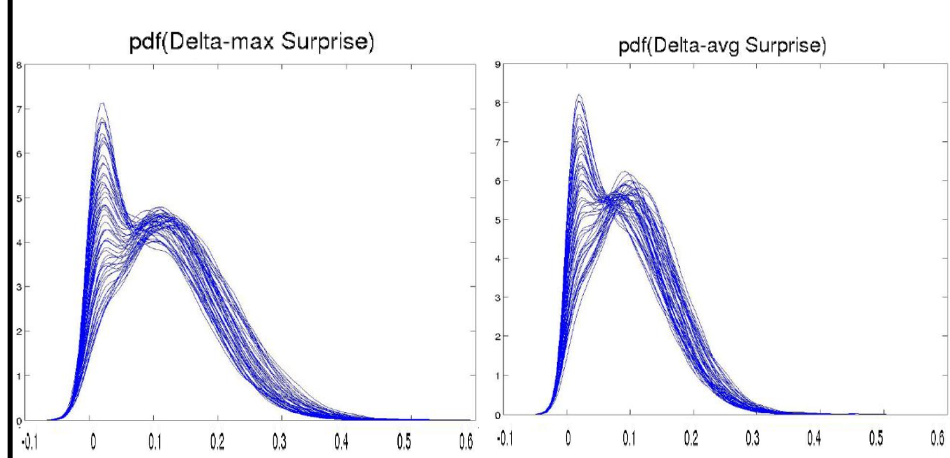
- Scenarios include expert decision similarity, agreement and disagreements.
  - Each scenario contains cases where there is / is not clipping in error distribution.
    - Each case covers conditions where noise causes/avoids label change in expert decisions.
- 100 observations for different experts outputs are given for an example 2-class scenario. Noise is distributed with stdev=1/15 and mean=0. Expert decision similarity, when surprise value of interest is 0, is analysed.

Case 1: No clipping due to noise
Condition: No label change

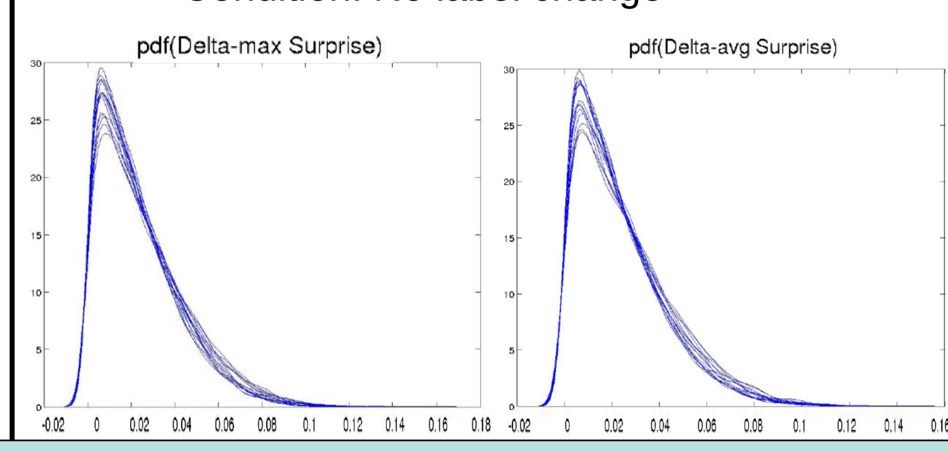




Condition: Label change



Case 2: Clipping due to noise
Condition: No label change



CONCLUSIONS

- Δ_{avg} gives more compact surprise distribution results than Δ_{max}^{dm}, when error is present.
- Label change during noise addition causes shifts in the surprise distribution for Δ_{max}^{dm} and Δ_{avg}. Shifts are more severe while using Δ_{max}^{dm}.
- BS is not effected by label change, however it has a range that cannot easily be thresholded for surprise detection, due to its divergence.
- Threshold value of 0.5 can be utilized for majority of the cases.

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) and Dstl.