Discrimination and its sensitivity in probabilistic networks

Silja Renooij and Linda C. van der Gaag

Contents

Context: applications for multiple-disorder diagnosis

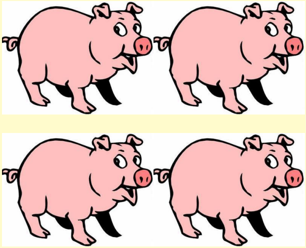
- the concept of evidence specific discrimination
- how to measure evidence specific discrimination
- how to study robustness of evidence specific discrimination
- conclusions and further research

Discrimination

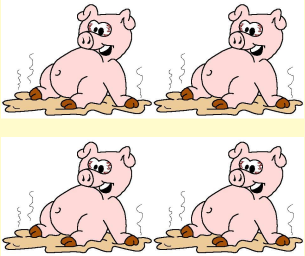
the ability to divide cases into competing classes

*discrimination **should** be good:*

Healthy:



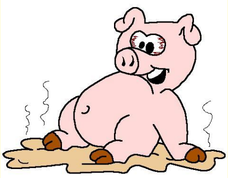
Not healthy:



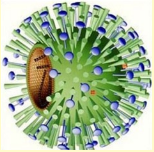
This is NOT what we are interested in here!

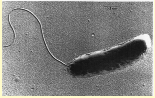
Evidence specific discrimination

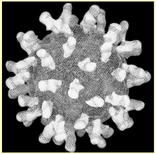
the ability to distinguish between multiple, not necessarily competing, “classes” *within* a single case

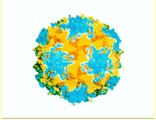


Classical Swine Fever and/or some other infection(s)?:



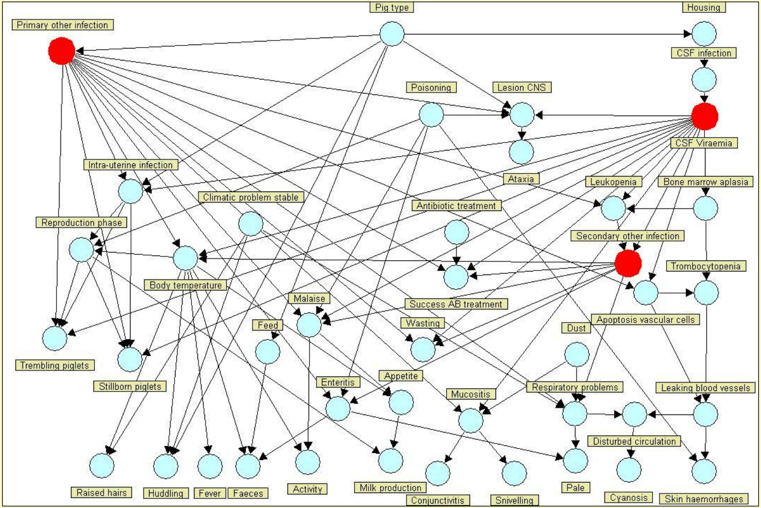






how *good is discrimination for this and similar cases?*

Motivation: early detection of CSF



Early detection of classical swine fever

We have multiple diagnostic nodes, since

- values are not mutually exclusive: csf can occur with other infections
- $\Pr(\mathrm{csf}) = 0.0000016$, but infections with similar symptoms are very common
- synergistic effects are crucial for diagnosis

Measuring evidence specific discrimination

We propose a measure of discrimination $d(a\,;b \mid e)$ between a and b in the context of evidence e, where a and b can be simple or compound values:

Examples:

- $d(\mathrm{POI} = \textit{gi}\,; \mathrm{POI} = \textit{ai} \mid \mathrm{Case}\,14)$
- $d(\mathrm{CSF} = \textit{csf}\,; \mathrm{POI} = \textit{gi} \mid \mathrm{Case}\,169)$
- $d(\mathrm{CSF} = \textit{csf}, \mathrm{POI} = \textit{ai}\,; \mathrm{CSF} = \textit{no-csf}, \mathrm{POI} = \textit{ai} \mid \mathrm{Case}\,304)$

Measuring evidence specific discrimination

We propose a measure of discrimination $d(a\,;b \mid e)$ between a and b in the context of evidence e:

- based on posterior probabilities: $\Pr(a \mid e)$ and $\Pr(b \mid e)$
- showing no discrimination when $\Pr(a \mid e) = \Pr(b \mid e)$
- giving maximum discrimination when $\Pr(a \mid e) = 1$ and $\Pr(b \mid e) = 0$, or vice-versa

Examples:

$$|p_a - p_b|, \quad \frac{|p_a - p_b|}{p_a + p_b}, \quad \left|\ln \frac{p_a}{p_b}\right|, \quad \left|\ln \frac{p_a/(1-p_a)}{p_b/(1-p_b)}\right|$$

where $p_a = \Pr(a \mid e)$ and $p_b = \Pr(b \mid e)$

Robustness of discrimination

"How robust is discrimination to inaccuracies in the network parameters"?

Idea:

- $d(a\,;b \mid e)$ is defined in terms of $\Pr(a \mid e)$ and $\Pr(b \mid e)$
- $f_a^e(x) = \Pr(a \mid e)(x) = \dfrac{\Pr(a, e)(x)}{\Pr(e)(x)}$ is the sensitivity function

  relating $\Pr(a \mid e)$ to a parameter x

  $\Longrightarrow$
- define $d(a\,;b \mid e)(x)$ in terms of $f_a^e(x)$ and $f_b^e(x)$ ...

Sensitivity functions for multiple nodes of interest

$$f_a^e(x) = \Pr(a \mid e)(x) = \frac{\Pr(a,e)(x)}{\Pr(e)(x)} = \frac{c_1 \cdot x + c_2}{c_3 \cdot x + c_4}$$

Known: constants can be efficiently established with standard inference algorithms, if a is a value for a single node of interest

For compound values, we now observe the following:

$$\Pr(a, b \mid e) = \Pr(a \mid b, e) \cdot \Pr(b \mid e) = \frac{\Pr(a,b,e)}{\Pr(b,e)} \cdot \frac{\Pr(b,e)}{\Pr(e)}$$

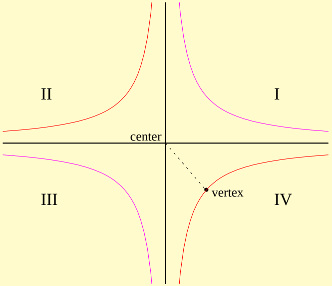
So

$$f_{ab}^e(x) = f_a^{be}(x) \cdot f_b^e(x) = \frac{c_1 \cdot x + c_2}{c_3 \cdot x + c_4} \cdot \frac{c_3 \cdot x + c_4}{c_5 \cdot x + c_6} = \frac{c_1 \cdot x + c_2}{c_5 \cdot x + c_6}$$

Back to robustness of discrimination

$d(a\,;b \mid e)(x)$ where a and b can be simple or compound values.

A sensitivity function is a fragment of a rectangular hyperbola branch:



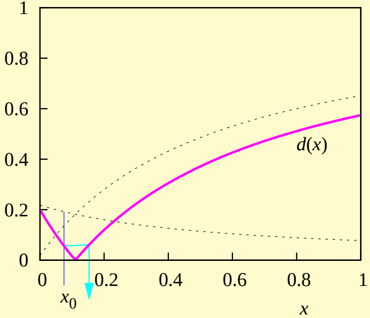
The same holds for $d(a;b \mid e)(x)$, depending on the definition of $d(a\,;b \mid e)$ used!

Examples:

$$f_a^e - f_b^e, \quad \frac{f_a^e - f_b^e}{f_a^e + f_b^e}, \quad \frac{f_a^e}{1 - f_a^e}$$

$\Longrightarrow$ knowing the functional form, constants for $d(a\,;b \mid e)(x)$ can be directly determined from those for $f_a^e(x)$ and $f_b^e(x)$.

Dynamics of discrimination



Example:

$$d(\mathrm{CSF} = \textit{csf}\,;\mathrm{POI} = \textit{gi} \mid 169)(x)$$
$$=$$
$$\left| f^{169}_{\textit{csf}}(x) - f^{169}_{\textit{gi}}(x) \right|$$

Given $d(a\,;b \mid e)(x)$, the following questions can be answered:

- for which value of x is discrimination maximised and what is this maximum value?
- for which value of x is discrimination minimised and what is this minimum value?
- for which value of x is the amount of discrimination the same as for x_0?

Conclusions and further research

- concept of evidence specific discrimination between simple or compound values
- properties and examples of discrimination measures
- sensitivity functions for compound values of interest
- use of sensitivity functions to study the robustness of discrimination

- when to use what measure?
- what amount of discrimination is acceptable or desirable?
- network-independent or evidence-dependent bounds on discrimination