DEPARTMENT OF SCIENCE, TECHNOLOGY,
ENGINEERING AND PUBLIC POLICY

UCL

How do public sector values enter today's public sector machine learning systems? (if at all!)

The Human Use of Machine Learning Workshop

European Centre for Living Technology, Venice 16/12/2016

Michael Veale
Department of Science, Technology, Engineering & Public Policy (UCL STEaPP)
University College London
m.veale@ucl.ac.uk / @mikarv

UCL ENGINEERING
Change the world

EPSRC
Engineering and Physical Sciences Research Council

STEaPP
Applied in Focus. Global in Reach.

current applications of ML in the public sector

current applications of ML in the public sector

anticipating

current applications of ML in the public sector

UCL

anticipating	crime hotspots

current applications of ML in the public sector

UCL

anticipating

- crime hotspots
- abusive households

current applications of ML in the public sector

UCL

anticipating

- crime hotspots
- abusive households
- food safety breaches

current applications of ML in the public sector

UCL

anticipating

- crime hotspots
- abusive households
- food safety breaches
- ‘solvability’ of crimes

current applications of ML in the public sector

UCL

anticipating

- crime hotspots
- abusive households
- food safety breaches
- ‘solvability’ of crimes
- firm insolvency

current applications of ML in the public sector

UCL

anticipating

- crime hotspots
- abusive households
- food safety breaches
- ‘solvability’ of crimes
- firm insolvency

detecting

current applications of ML in the public sector

UCL

anticipating

- crime hotspots
- abusive households
- food safety breaches
- ‘solvability’ of crimes
- firm insolvency

detecting

- fraudulent tax returns

current applications of ML in the public sector

UCL

anticipating

- crime hotspots
- abusive households
- food safety breaches
- ‘solvability’ of crimes
- firm insolvency

detecting

- fraudulent tax returns
- incorrectly coded crime records

current applications of ML in the public sector

UCL

anticipating

- crime hotspots
- abusive households
- food safety breaches
- ‘solvability’ of crimes
- firm insolvency

detecting

- fraudulent tax returns
- incorrectly coded crime records
- mobile homes for address registers

current applications of ML in the public sector

UCL

anticipating

- crime hotspots
- abusive households
- food safety breaches
- ‘solvability’ of crimes
- firm insolvency

detecting

- fraudulent tax returns
- incorrectly coded crime records
- mobile homes for address registers
- changes in stats between censuses

methodology

Q: How do issues around ethics/responsibility emerge in practice, and how do public sectors/contractors perceive and cope with them?

Interviews undertaken with 30+ actors in public sector machine learning in five countries —

- 'screen level' bureaucrats
- 'system level' bureaucrats/responsible civil servants
- technologists/technology brokers

what are public sector values?

UCL

Big debate in public administration literature. Include:

robustness usability legality

upskilling productivity dialogue

advocacy–neutrality innovation equity

accountability competition–cooperation openness–secrecy

what are public sector values?

UCL

Let’s zoom in on a few of these for now

robustness usability legality

upskilling productivity dialogue

advocacy–neutrality innovation equity

accountability competition–cooperation openness–secrecy

equity in theory: **preventing discrimination**

what kind of discrimination?	**direct** (use of protected characteristics) **indirect** (use of correlated characteristics) **both** (mix)
what kind of prevention?	**preprocessing** (massage the data) **inprocessing** (change the learning logic) **postprocessing** (alter the learned model)

For more, see Kamiran, F. et al. (2012). Techniques for Discrimination-Free Predictive Models . doi: 10.1007/978-3-642-30487-3_12

equity in practice: **fairness–performance tradeoffs**

UCL

> We decided in the end to remove sensitive variables such as race and gender. Some people will argue that we might be being implicitly biased through other variables. **I've even heard that we should strip out location entirely**. One thing you can do is you can **make a model with and without the sensitive variables and see what lift you get in comparison**. That way you can **make it clearer what the options are** and allow the clients to trade them off.

— Contractor who led a predictive policing project for a global city

equity in practice: **scientific advice for correlations**

UCL

> Whether a **child is deaf or disabled is empirically linked to abuse, according to [NGO] research**. But of course [local governments] are also aware they don't want parents singled out as potential abusers simply because they have a disabled child. Poverty is another correlating factor — for example, free school meals by virtue of lack of ability to pay.

— Police chief leading a anticipatory child protection project

equity in practice: **ML can itself unearth unfairness**

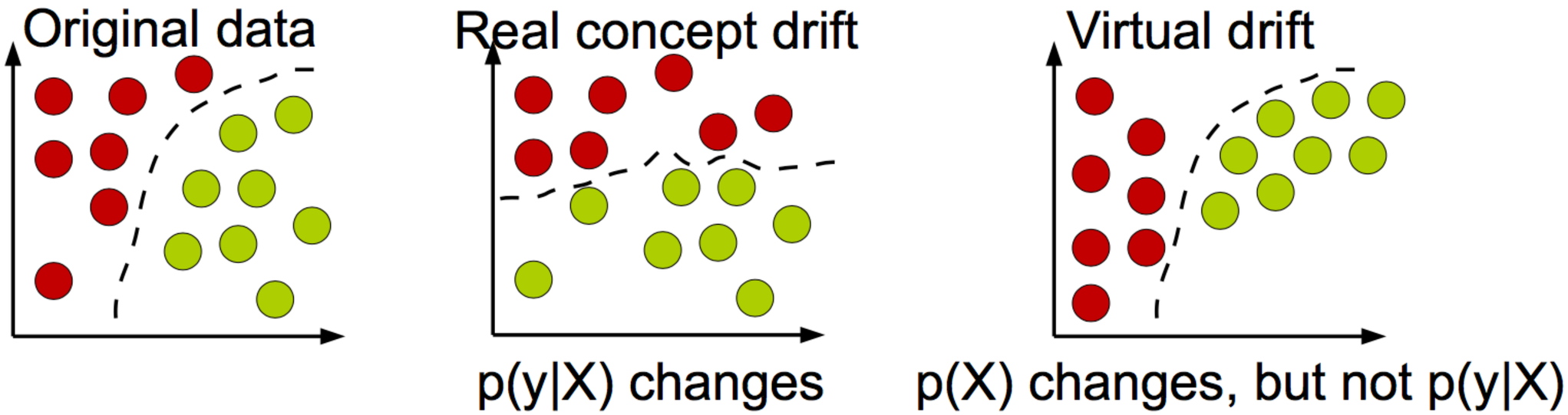
UCL

> Individual judgement also come with their own biases. **We will surely find things that are uncomfortable, unpleasant, even shocking, and we'll have to face up to those and be happy we discovered them**. This is realistically likely to be what [policy partner] is scared of, y'know — *oh, shucks! what will this algorithm unearth?!*

— NGO partner on an anticipatory child protection project

robustness in theory: **concept drift**

1. link between output and input changes [real concept drift]
2. distribution of inputs change [virtual concept drift]



For more, see Gama, J. et al. (2013). A Survey on Concept Drift Adaptation. doi: 10.1145/0000000.0000000 [above diagram from paper]

robustness in practice: **challenges of legacy systems**

UCL

> Historically, we have hard-coded equations into operational systems, with the weights on the regression that we determined. Input variables could then be added manually by staff in prisons, which was time consuming. **Hardcoding creates two main consequences. The first is that updating the model costs a fortune. The second, which follows from the first, is that we don't update often.**

— Public servant building models for a national prison system

robustness in practice: **data collection loops**

UCL

> Thankfully we barely have any reports of human trafficking. But someone at intel got a tip-off and looked into cases of human trafficking at car washes, because we hadn't really investigated those much. **But now when we try to model human trafficking we only see human trafficking being predicted at car washes, which suddenly seem very high risk**. So because of increased intel we've essentially produced models that tell us where car washes are. This kind of loop is hard to explain to those higher up.

— In-house police department machine learning modeller

robustness in practice: **the world speaks back**

> The highest probability assessments are on the mark, but actual deployment causes displacement, dispersion and diffusion, and that **throws the algorithm into a loop** [...] **as you deploy resources, displacement and dispersal goes through the roof** [...] In the first four weeks of trialling it out, **the probability of being correct just tanked**

— Police head of analytics for a major world city

accountability in theory: **interpretability**

?

decompositional

make a more
interpretable algorithm

regression
decision trees

pedagogical

wrap an uninterpretable algorithm
with a simpler one to estimate
its logics

LIME [arxiv:1602.04938]
rule extraction

accountability in practice: **interpretation's pitfalls**

UCL

> To explain these models we talk about the target parameter and the population, rather than the explanation of individuals. The target parameter is what we are trying to find — the development of debts, bankruptcy in six months. The target population is what we are looking for: for example, businesses with minor problems. **We only give the auditors [these], not an individual risk profile or risk indicators [...] in case they investigate according to them.**

— Public servant responsible for ML at a national tax office

accountability in practice: **humans-in-the-loop**

UCL

> We ask local officers, intelligence officers, to look at the regions of the [predictive project name] maps which have high predictions of crimes. They are the people who file or read all the local reports that are made, as well as other sources of information about those areas. **They might say they know something about the offender for a string of burglaries, or they might say that a high risk building is no longer at such high risk of burglary because they local government just arranged all the locks in that building to be changed.**

— Police lead on a national predictive policing project

concluding points

- Technical solutions don't fit neatly into the needs of different actors.
- Feedback is especially powerful in high stakes environments.
- External knowledge/expert advice currently filling in the hole from the lack of fairness technologies
- Tradeoffs within whole sociotechnical systems, not within narrow well-defined mathematical problems.

thanks!

Q?

@mikarv
m.veale@ucl.ac.uk

robustness in practice: **individuals matter**

> There's one woman who calls in whenever her kid is out after 10pm. She then calls back about 30 mins or so later to say that everything is fine, or we follow up with her. **But then it looks like in the model that kids always go missing at 10pm**, which obviously is a bit misleading. In the end I had to manually remove her from the model to remove the spurious pattern.
>
> **— In-house police department machine learning modeller**

accountability in practice: **occam's razor**

UCL

> [We] built something huge, 18,000 variables initially. We then narrowed these to about 200, then to about 20. Keeping one year of records out, we tested quarter by quarter to refine and build the model and choose these variables. Then we could bring in more complex models, like Random Forests, and use those in addition. **You're familiar with the term Occam's Razor? We honed it down in the end to eight variables, because it's important to see how it works, we believe.**
>
> **— Contractor who led a predictive policing project for a global city**

accountability in practice: **humans-in-the-loop**

UCL

> We also have weekly meeting with all the officers, leadership, management, patrol and so on, and the intelligence officers are the core of this meeting. **There, he or she presents what they think is going on in this map, and what should or could be done about it.**

— Former police lead on a national predictive policing project

accountability in practice: **openness vs gaming**

> In compliance models we don't give many details. We might say we are interested in sectors or size, and perhaps share the weights with one or two key people. With regards to this, **we're primarily concerned that if the model weights were public, their usefulness might diminish.**
>
> **— Head of Analysis at a national tax office**

equity in practice: **measuring v managing**

UCL

> Race is very predictive of reoffending [...] we don't include race in our predictive models. [...] **we are aware that we are using conviction as the proxy variable for offending** [...] you can get into cycles looking at certain races which might have a higher chance of being convicted [...] you're **building systems and catching people not based on the outcome, but on the proxy outcomes**.

— Public servant building models for a national prison system