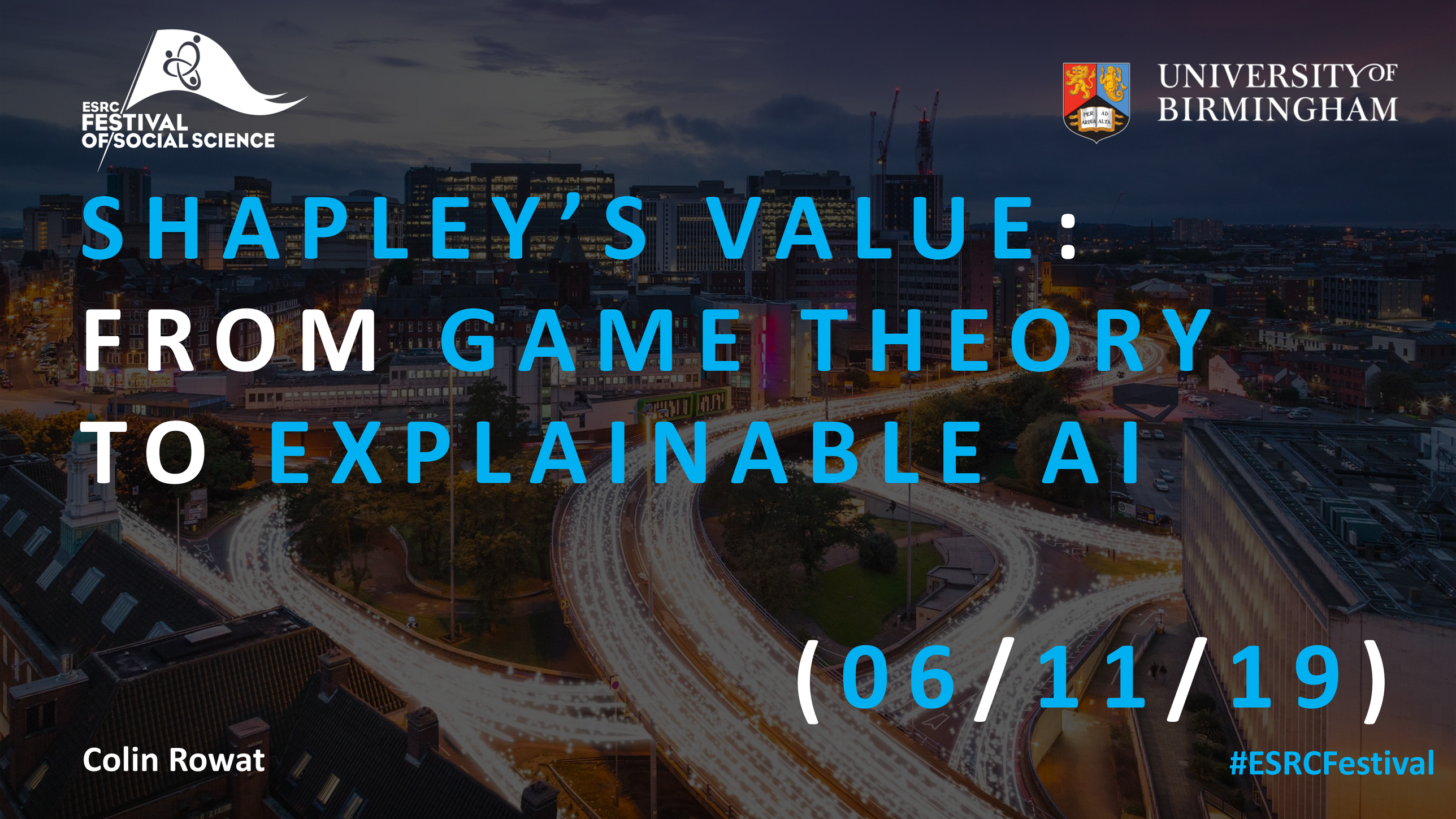

ESRC
FESTIVAL
OF SOCIAL SCIENCE
UNIVERSITY OF
BIRMINGHAM
SHAPLEY'S VALUE:
FROM GAME THEORY
TO EXPLAINABLE AI
(06/11/19)
Colin Rowat
#ESRCFestival

von Neumann matches gloves

- Z has a left glove
- K & S each have a right glove
- a coalition only scores if it has a pair:
  v(S,Z) = v(K,Z) = v(S,K,Z) = 1
- a von Neumann-Morgenstern solution:
  - Z has a veto
  - K & S together have a veto
  - any outcome where S,K share equally

yuck!

1. many outcomes in the solution
2. many asymmetric solutions
3. solutions might not exist
4. solutions might be really ugly

Colin Rowat

#ESRCFestival

ESRC FESTIVAL OF SOCIAL SCIENCE

enter the Shapley value, I

UNIVERSITY OF BIRMINGHAM

what *should* each player receive?

- average over value it adds to coalitions
  - p = ⅓: K joins empty coalition; adds 0
  - p = ⅓ × ½: K joins S; adds 0
  - p = ⅓ × ½: K joins Z; adds 1
  - p = ⅓: K joins S, Z; adds 0
  - thus, K's Shapley value is ⅙

yum!

1. a single, unique result, (⅙, ⅙,⅔)
2. always exists
3. easy to calculate
4. symmetric

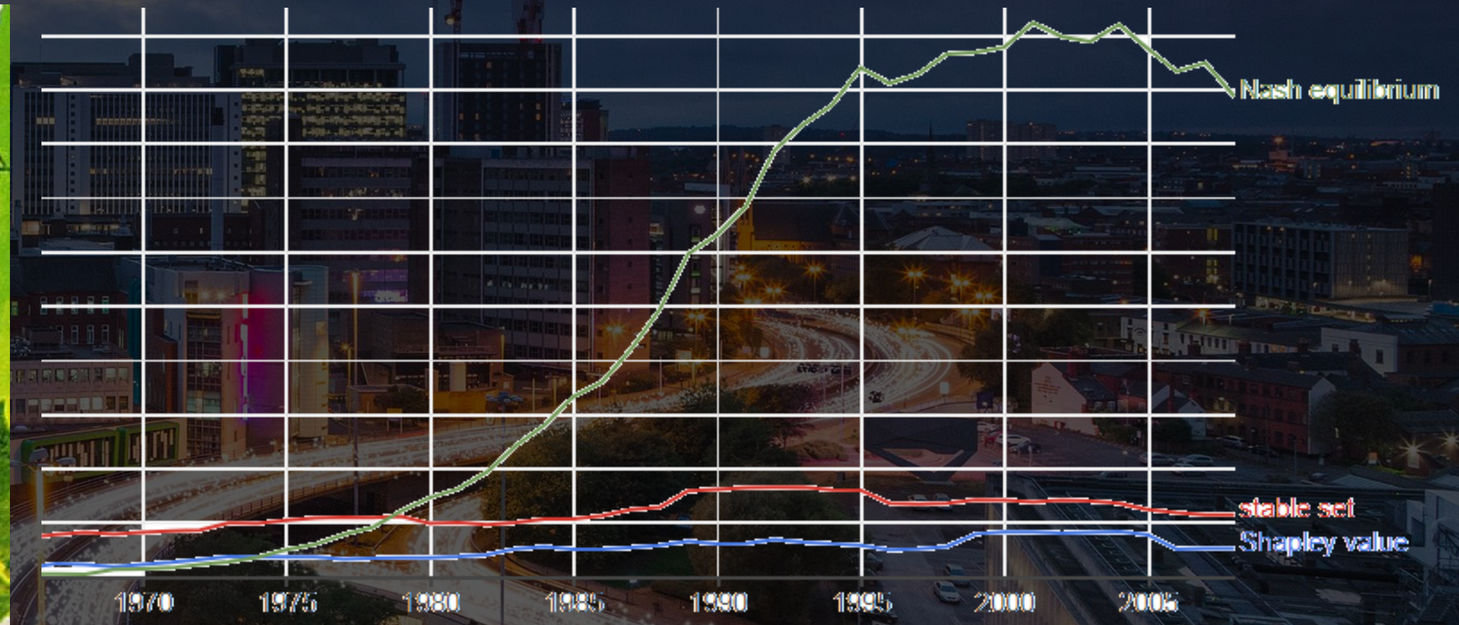
Colin Rowat

#ESRCFestival

ESRC FESTIVAL OF SOCIAL SCIENCE

exit von Neumann, Shapley

UNIVERSITY OF BIRMINGHAM



Why does Nash's equilibrium take over?

1. handles information better
2. constructive: *how* do we get an outcome?

Colin Rowat

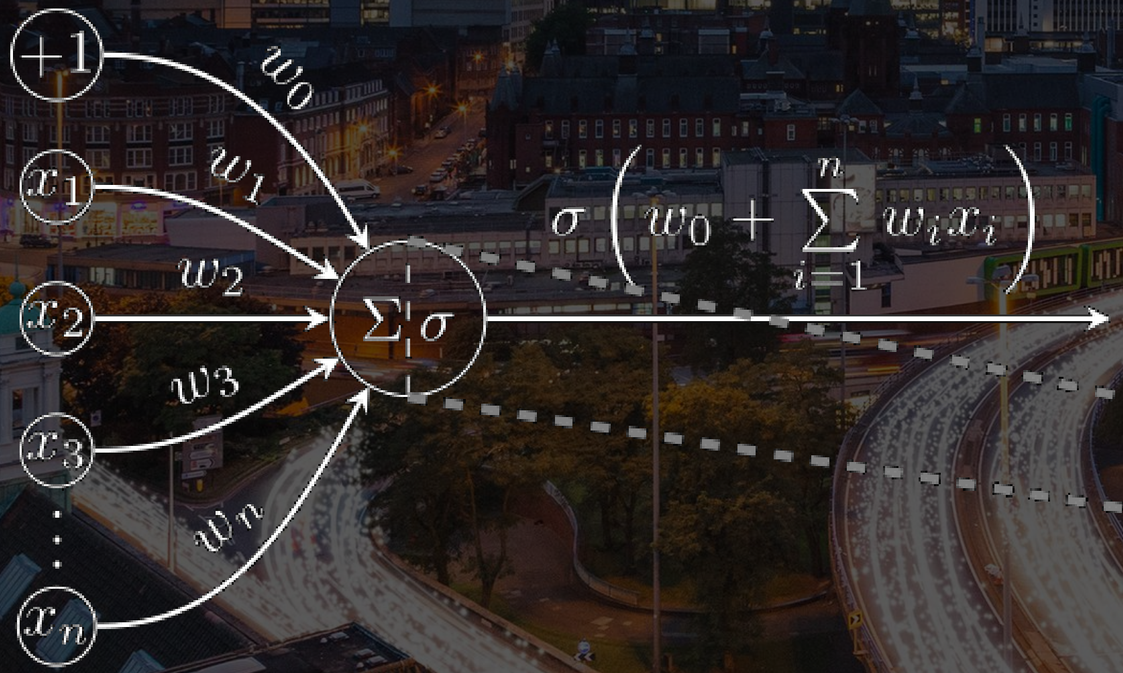
asterix.fandom.com

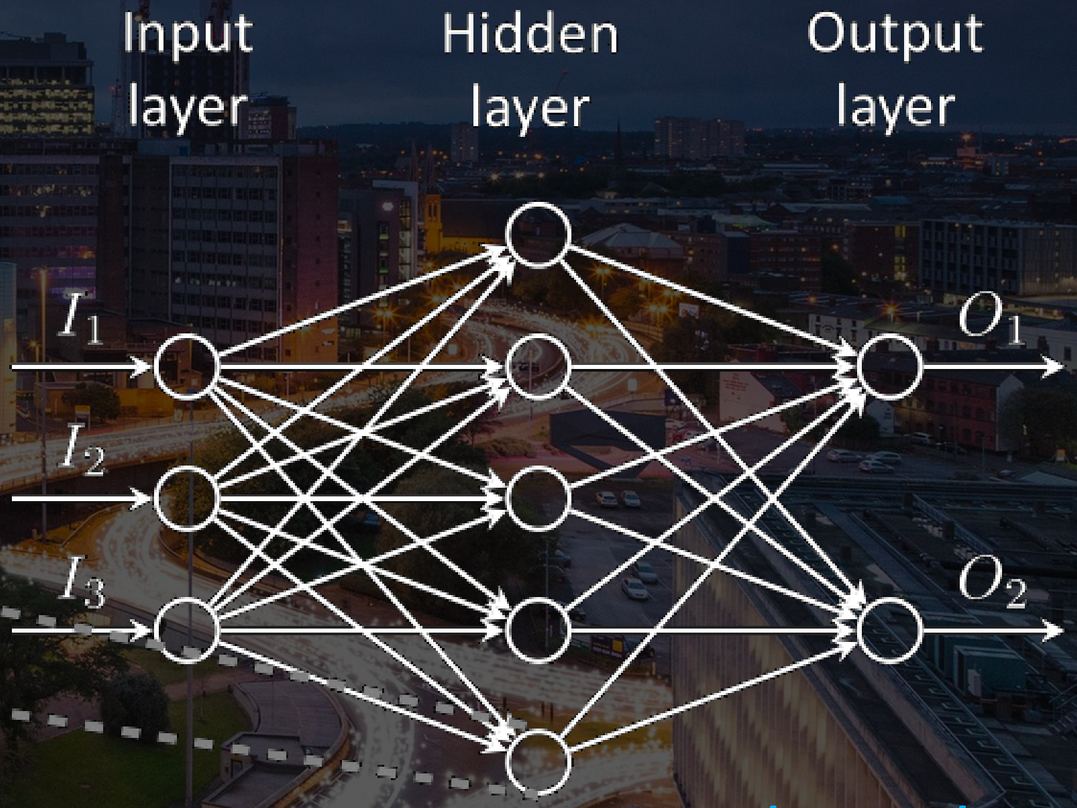
#ESRCFestival

ESRC FESTIVAL OF SOCIAL SCIENCE

rise of the machines

UNIVERSITY OF BIRMINGHAM





github.com/PetarV-/TikZ

When an neural network/AI system makes a prediction, how do we explain it?

Colin Rowat

#ESRCFestival

ESRC FESTIVAL OF SOCIAL SCIENCE

enter the Shapley value, II

UNIVERSITY OF BIRMINGHAM

APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY
Appl. Stochastic Models Bus. Ind., 2001; **17**:319–330 (DOI: 10.1002/asmb.446)

Analysis of regression in game theory approach

Stan Lipovetsky*,† and Michael Conklin

SUMMARY

Working with multiple regression analysis a researcher usually wants to know a comparative importance of predictors in the model. However, the analysis can be made difficult because of multicollinearity among regressors, which produces biased coefficients and negative inputs to multiple determination from presumably useful regressors. To solve this problem we apply a tool from the co-operative games theory, the Shapley Value imputation. We demonstrate the theoretical and practical advantages of the Shapley Value and show that it provides consistent results in the presence of multicollinearity.

Knowl Inf Syst (2014) 41:647–665
DOI 10.1007/s10115-013-0679-x

Explaining prediction models and individual predictions with feature contributions

Erik Štrumbelj · Igor Kononenko

This leads to the following explicit definition (see [28] for proof):

$$\varphi_i(x) = \sum_{Q \subseteq S \setminus \{i\}} \frac{|Q|!(|S| - |Q| - 1)!}{|S|!} (\Delta_{Q \cup \{i\}}(x) - \Delta_Q(x)). \quad (9)$$

Equation (9) is equivalent to the Shapley value [25], a concept from coalitional game theory. In a coalitional game, it is usually assumed that n players form a grand coalition

Colin Rowat

#ESRCFestival

how did Shapley get here?

players ⇒ features

- **e.g. K ⇒ e.g. `kids allowed' …**

in the coalition ⇒ specific value of feature
out of the coalition ⇒ average value

e.g. $f(\overline{K},S_i,Z_i)$ **prediction given average K, specific values of S and Z**

'value added' ⇒ the prediction's change

$f(K,S,Z) - f(\overline{K},S,Z)$

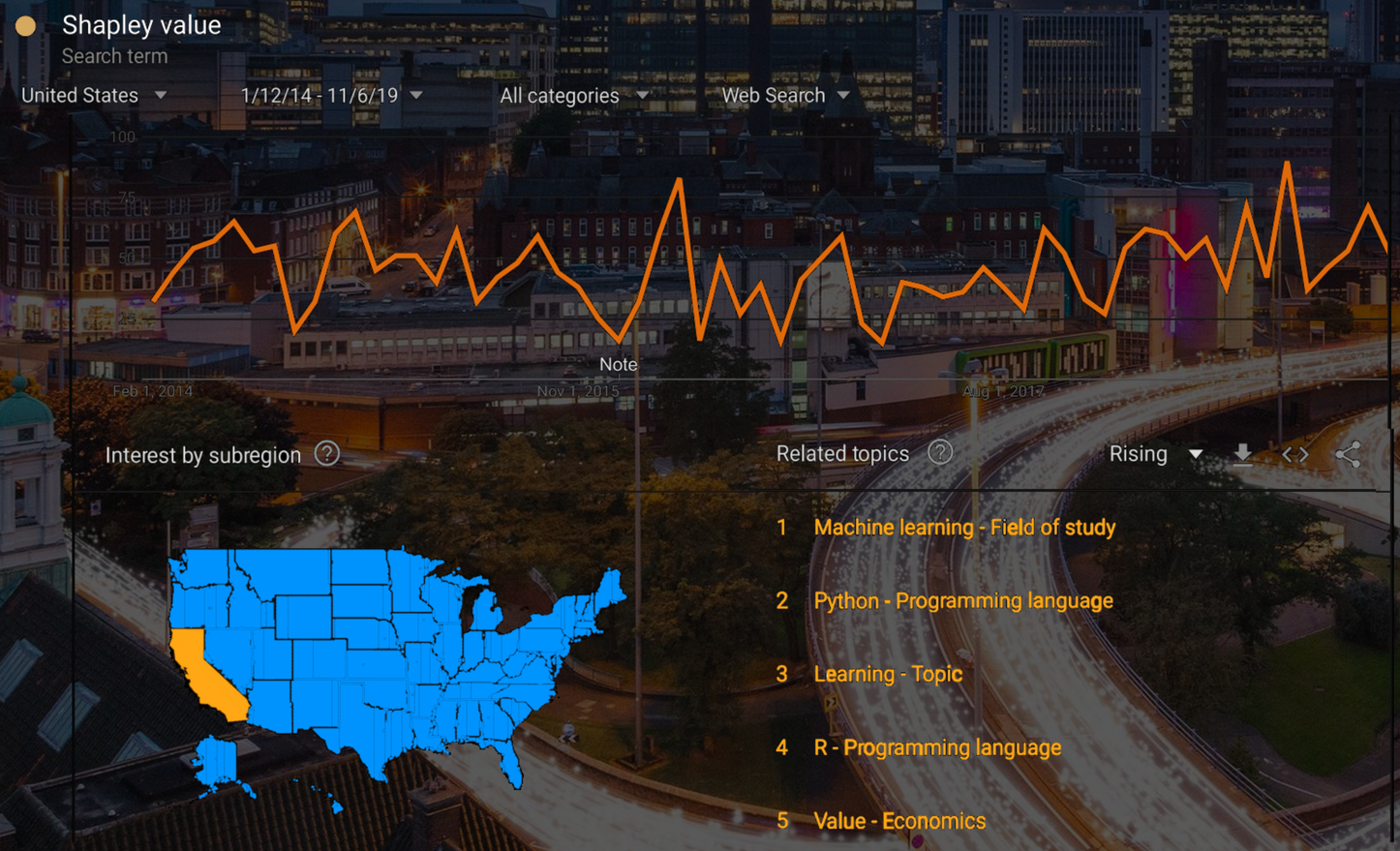
Colin Rowat

#ESRCFestival

ESRC FESTIVAL OF SOCIAL SCIENCE

Shapley redux?

UNIVERSITY OF BIRMINGHAM



The Shapley value might be the **only method** to deliver a full explanation. In situations where the law requires explainability – like EU's "right to explanations" – the Shapley value might be the **only legally compliant** method, because it is based on a solid theory and distributes the effects fairly. … The Shapley value is the **only explanation** method with a solid theory. (Molnar, 03/11/19)

Colin Rowat

#ESRCFestival

ESRC FESTIVAL OF SOCIAL SCIENCE

what if we break symmetry?

UNIVERSITY OF BIRMINGHAM

don't assume equal arrival probabilities

- is this just obscure theory?

Probabilistic values for **games**

RJ **Weber** - The Shapley **Value**. Essays in Honor of Lloyd S …, **1988** - books.google.com

The study of methods for measuring the "value" of playing a particular role in an n-person game is motivated by several considerations. One is to determine an equitable distribution of the wealth available to the players through their participation in the game. Another is to help an individual assess his prospects from participation in the game. When a method of valuation is used to determine equitable distribu-tions, a natural defining property is "efficiency": The sum of the individual values should equal the total payoff achieved through …

☆ Cited by 546 Related articles »

Variations on the Shapley value

D Monderer, D Samet - Handbook of game theory with economic …, 2002 - Elsevier

This survey captures the main contributions in the area described by the title that were published up to 1997.(Unfortunately, it does not capture all of them.) The variations that are the subject of this chapter are those axiomatically characterized solutions which are …

☆ Cited by 85 Related articles »

Colin Rowat

#ESRCFestival

ESRC FESTIVAL OF SOCIAL SCIENCE

correlation isn't causation

UNIVERSITY OF BIRMINGHAM

e.g. hotel occupancy data

- price ∝ occupancy rates
- hike prices to increase profits?

use DAGs to set Shapley arrival order?

- let causes arrive before effects

Colin Rowat

#ESRCFestival

ESRC FESTIVAL OF SOCIAL SCIENCE

explaining high earnings

UNIVERSITY OF BIRMINGHAM

causal ancestors

- age, sex, native country, race

causal descendants

- marital-status, education…

some values seen as less important

- e.g. marital-status

others seen as more important

- e.g. sex (largest), race

Frye, Feige & Rowat (2019)

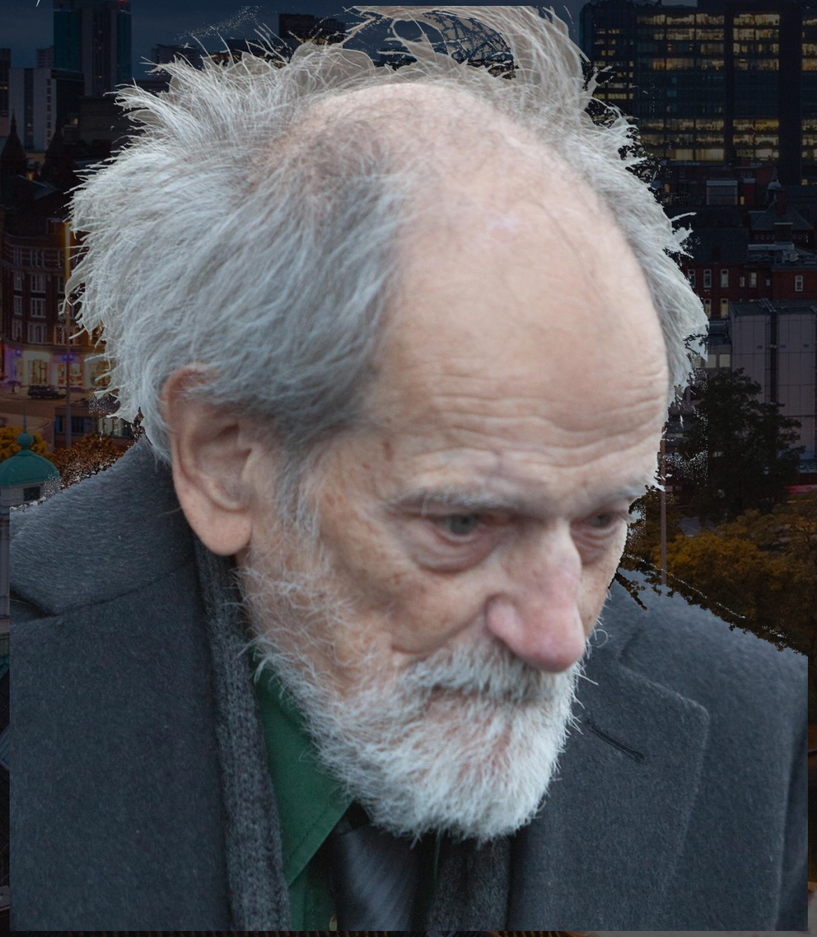
Colin Rowat

#ESRCFestival

ESRC FESTIVAL OF SOCIAL SCIENCE

conclusions

UNIVERSITY OF BIRMINGHAM



Wikipedia Commons

Colin Rowat

… about ideas and tools

- sometimes make unexpected leaps

… about explainable AI/ML

1. more sensible explanations
2. what are the right axioms?
3. single number search misguided?
   e.g. Wasserstein, Schirm, Lazar (2019) on $p < 0.05$
4. what about allowing joint arrival?

#ESRCFestival