Building Models from Big Data using Complex Systems Biases: Experience and Examples

Stan Matwin, CRC

Faculty of Computer Science
Dalhousie University
Halifax, NS, Canada

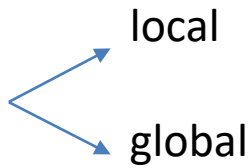
ICSP, Lucca, February, 2017

Roadmap

- The problem
- The idea
- Early experience

Critiques of Big Data

- Model interpretability [Guestrin 2016]



- Correlation vs causality
- Patterns/rules vs serendipity/nuggets [Helbing, Ratti 16]

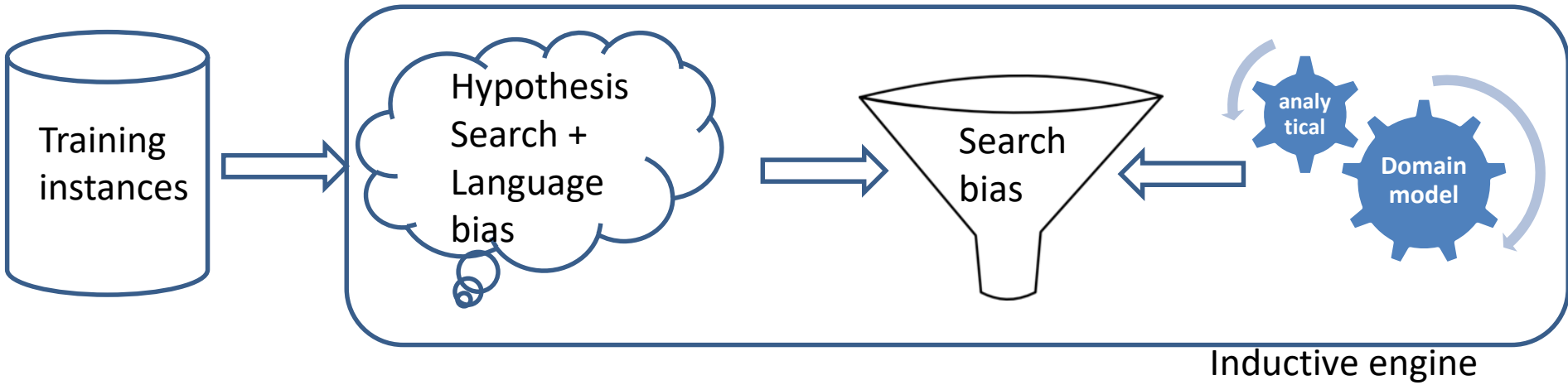
Causality and discovered knowledge

model	statistical/ correlational	causal
expected	most of current methods	**probabilistic** models: some Bayesian approaches [Perl, Scholkopf; Jensen]; **intensional** models: this proposal
serendipitous	outliers/ anomaly detection	**?**

Idea

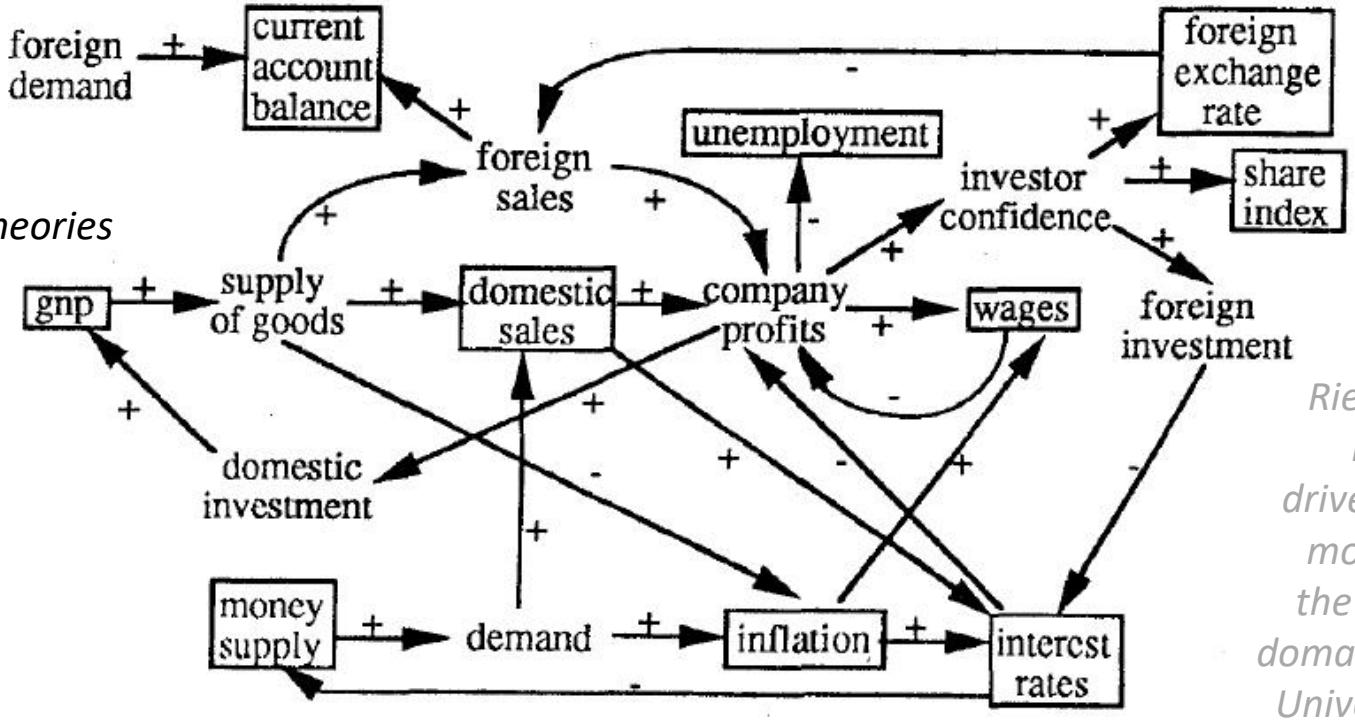
Background knowledge

- Models are built using Machine Learning techniques



Example

Clark, P., Matwin, S. Learning Domain Theories using Abstract Background Knowledge, ICML 93



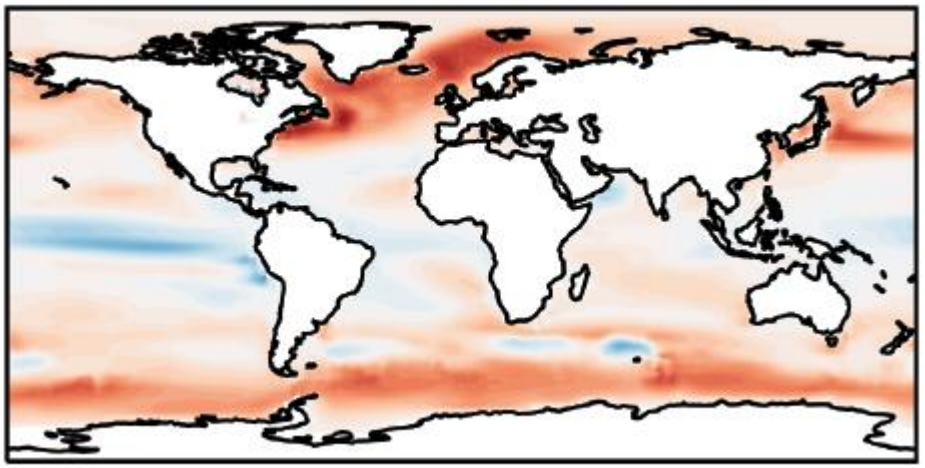
Spohrer, J., Riesbeck, Ch., Reasoning-driven memory modication in the economics domain. TR, Yale University, May 1984.

Only rules consistent with the QM are explored by inductive search

"If inflation grows
Then GNP decreases"

Planned Oceanographic application

- CO_2 sink:



- Data: observations of the surface ocean pCO_2 highly heterogeneously distributed in time and space
- SOM map model has been built [Landschutzer et al 2013]
- Such model could be “driven” by the prevailing time-series based climate model (“North Atlantic Oscillation”)

Conclusion

- Some interesting challenges for the “big data” paradigm
- Important for the acceptance of the data analysis results
- Initial ideas
- Lots of future work potential


Q
&
A