Autocoding the Survey of Occupational Injuries and Illnesses

Alexander Measure

Survey of Occupational Injuries and Illnesses

Example Narrative

Job title: sanitation worker

What was the employee doing just before the incident?
mopping floor in gym

What happened?
slipped on water on floor and fell

What part of the body was affected?
fractured right arm

What object directly harmed the employee?
wet floor

Codes Assigned

Occup: 37-2011 (Janitor)

Nature: 111 (Fracture)

Part: 420 (Arm)

Event: 422 (Fall, slipping)

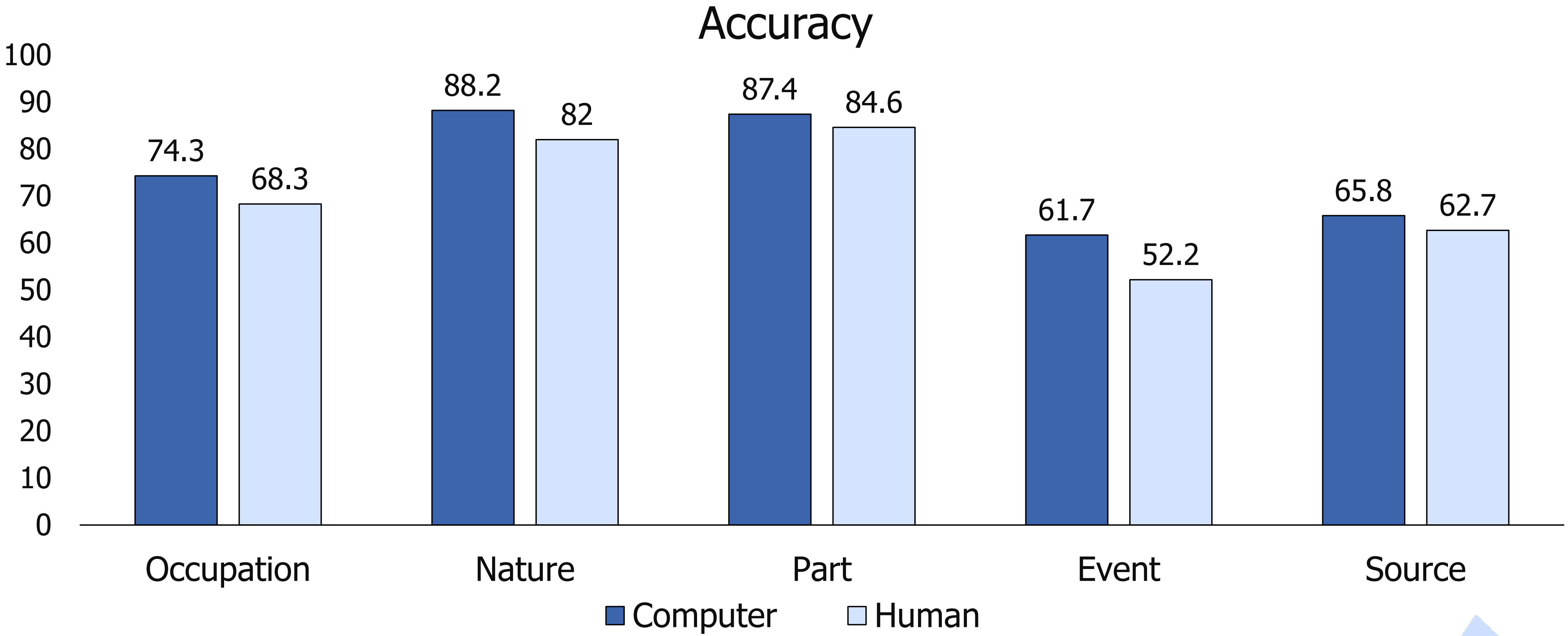
Source: 6620 (Floor)

Secondary: 9521(Water)

Supervised Machine Learning

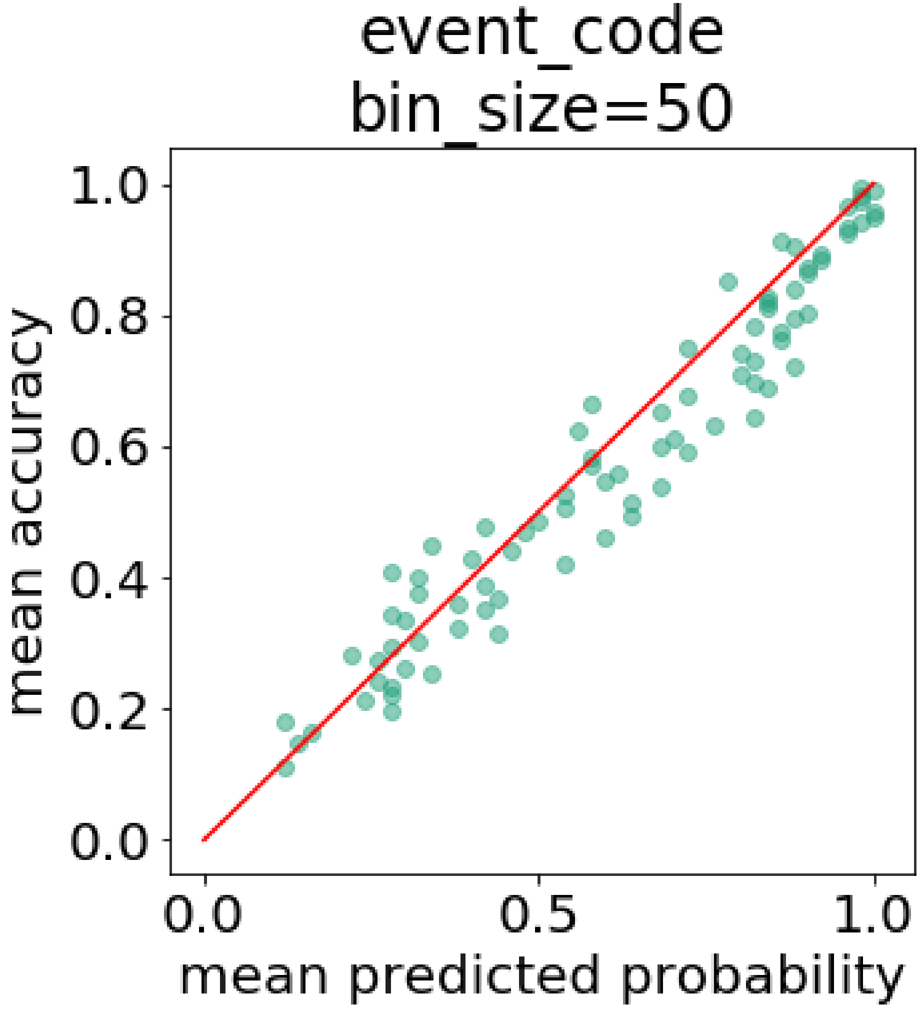
1. Data (training, validation, and test)
2. Determine inputs and numeric representation
3. Choose a learning algorithm
4. Fit to training data, evaluate on validation
5. Modify and repeat
6. At the very end, evaluate on test

Machine Learning vs. Manual Process



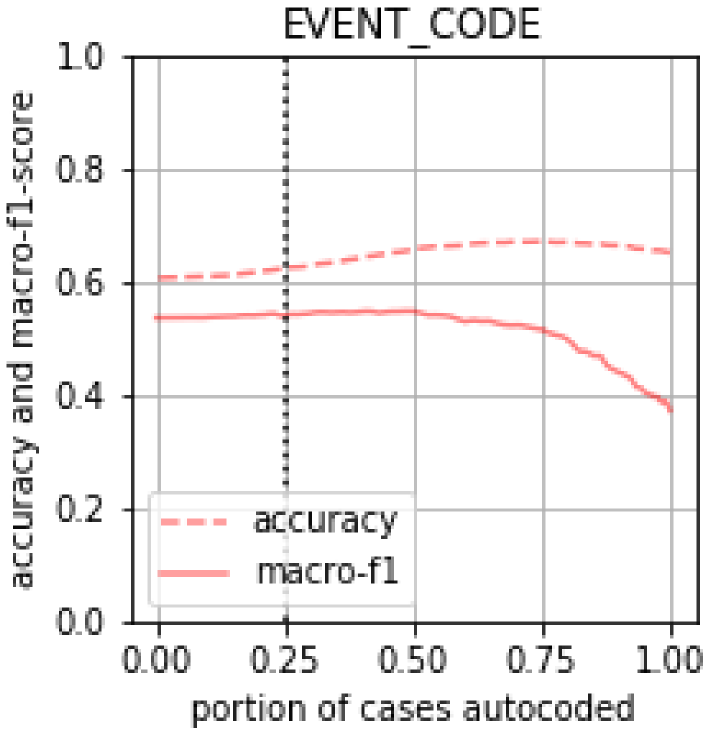
The benefits of probabilistic models

- Predicted Prob ≈ True Prob
  - It mostly knows what it doesn't know
  - Maybe a human knows?

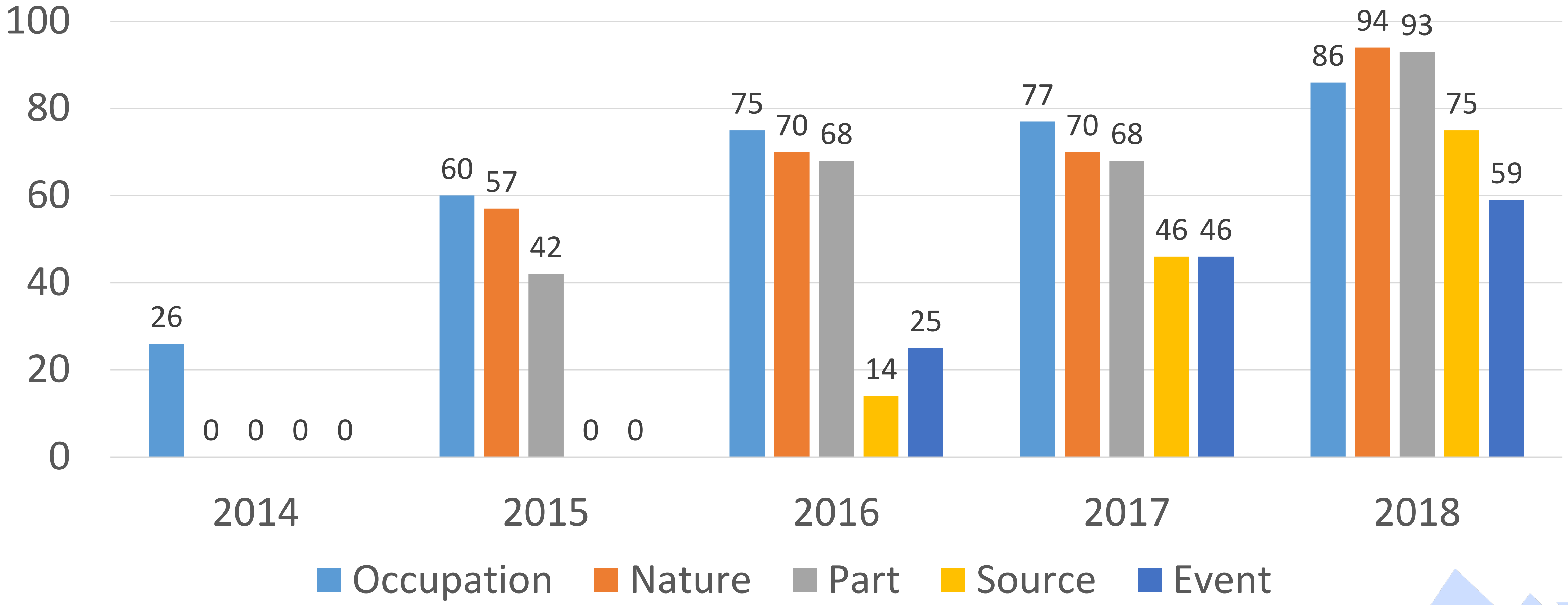


Maximizing Quality by Simulating Possibilities

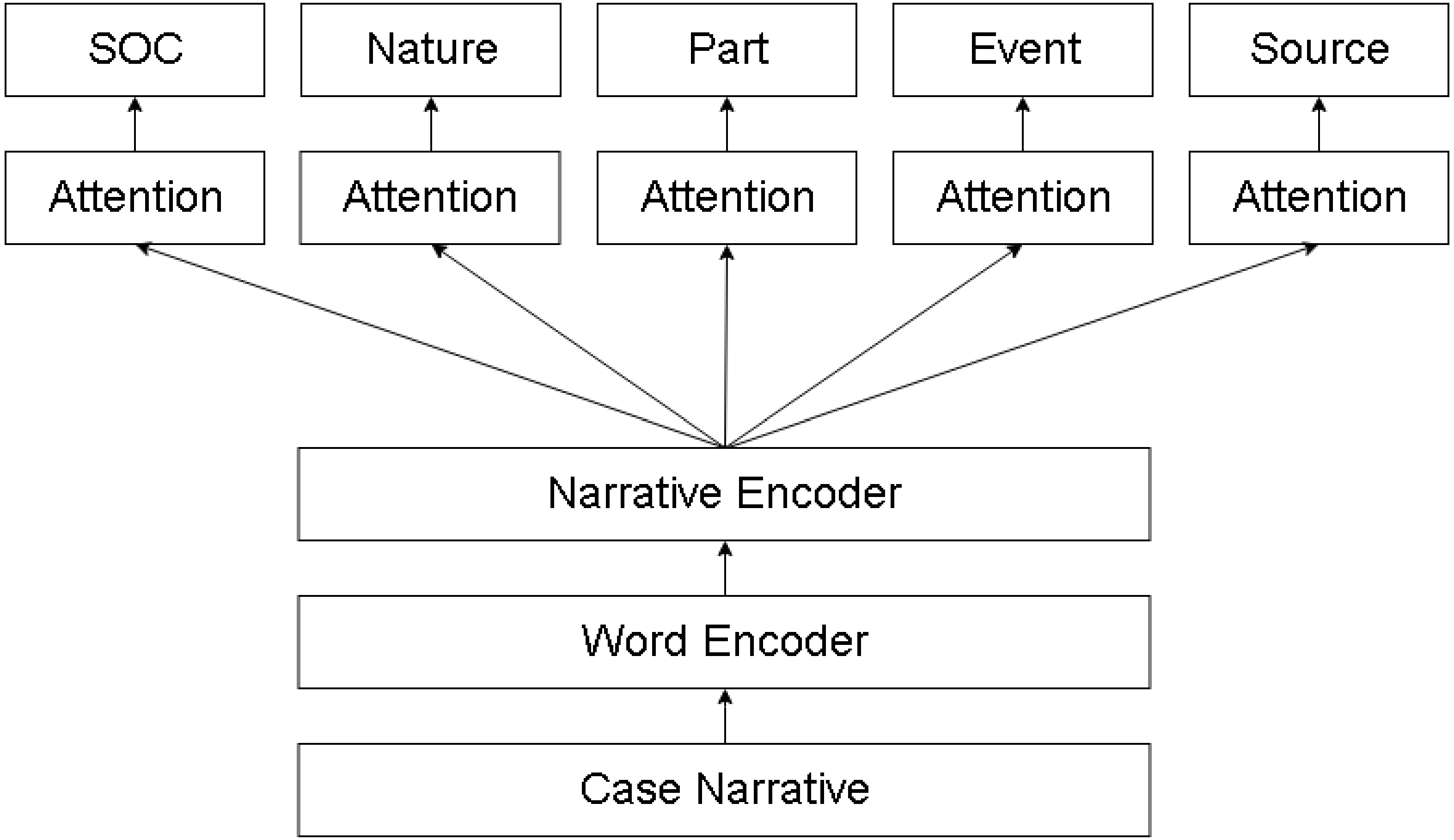
- Gold + Human + Computer codes allows simulation



% of codes automatically assigned to SOII



The Neural Network Autocoder



SOC

Nature

Part

Event

Source

Attention

Attention

Attention

Attention

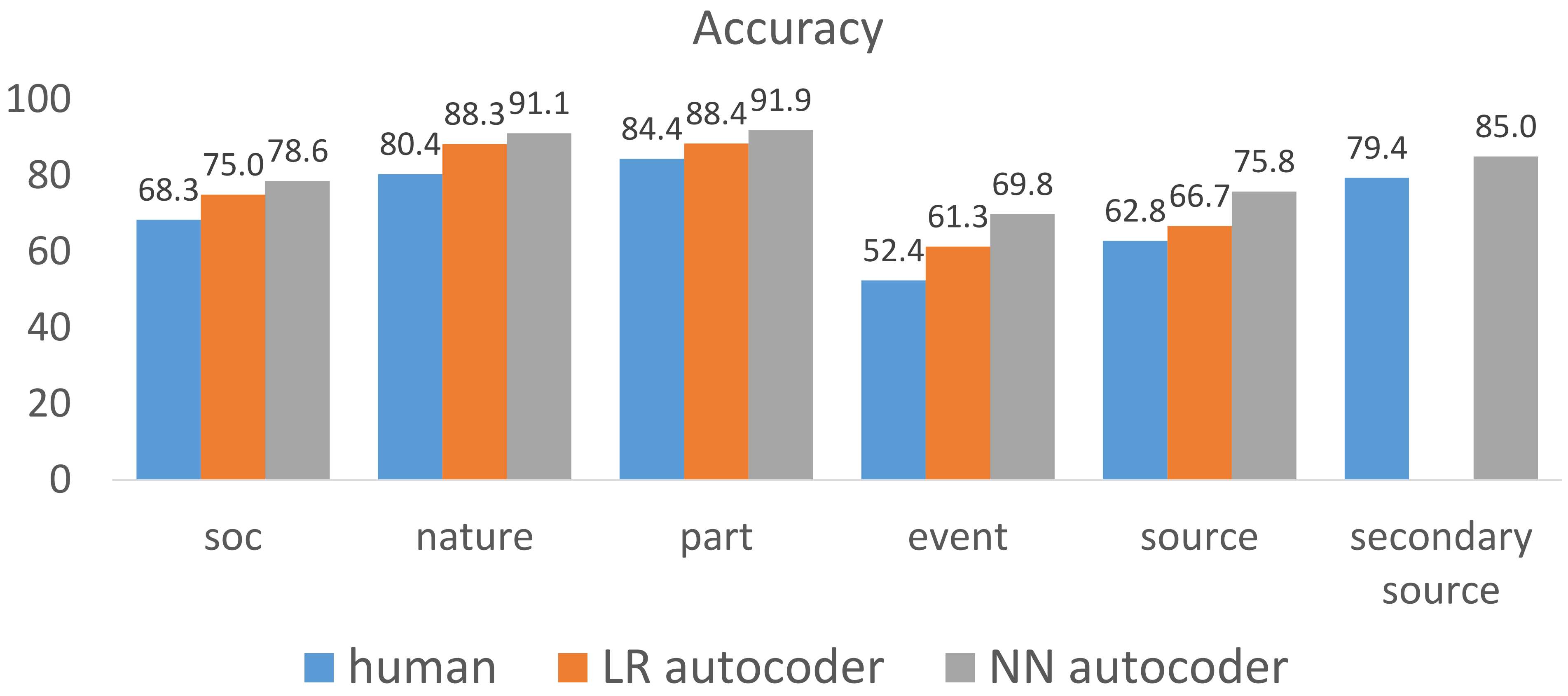
Attention

Narrative Encoder

Word Encoder

Case Narrative

Did it work?


Accuracy
100
80
60
40
20
0
68.3
75.0
78.6
80.4
88.3
91.1
84.4
88.4
91.9
52.4
61.3
69.8
62.8
66.7
75.8
79.4
85.0
soc
nature
part
event
source
secondary source
human
LR autocoder
NN autocoder

Things I wish someone had told me

- Gold standard
  - Not optional if you care about quality and replacing an existing process
  - It **must be blind** (reviewers 6x more likely to keep codes they see)
- Not that hard to create
  - Find an expert (or 2)
  - Ask them to recode your test set (without access to original codes)
  - Bigger is better but even 500 will get you a 95% CI +/- 4.5% accuracy

Things I wish someone had told me

- You’re not done once it’s in production
  - Machine learning models need monitoring and maintenance
  - Neither is trivial
- Approach that’s worked best so far
  - Hold back a “sample” for humans to code
  - Then recode with experts, and add to gold standard
  - Allows updating of human / computer accuracy metrics so you can maintain right mix

Things I wish someone had told me

- Don’t spend a lot of time trying every preprocessing, feature, and model possible
  - Most were designed for something else
  - Most don’t matter

- My best model and feature ideas always came from looking carefully at the errors the model was making and working out why that would happen.

What’s next?

- State-of-the-art NLP continues to advance rapidly
  - Transfer learning with pretrained models
  - BERT, RoBERTa, ALBERT
- Sharing models with the public
  - Differential privacy works here too!
- Training staff

Additional Resources

- Tutorials
  - Logistic Regression

    https://github.com/ameasure/autocoding-class/blob/master/machine_learning.ipynb
  - Neural Networks

    https://colab.research.google.com/drive/1g3MVMCLOYshI_gaqMkDDj9gtG7yQQxib?ts=5c98e613
- Papers
  - https://www.bls.gov/osmr/pdf/st140040.pdf
  - https://www.bls.gov/iif/deep-neural-networks.pdf
    - Code: https://github.com/USDepartmentofLabor/soii_neural_autocoder

Contact Information

Alexander Measure
Economist
Office of Safety and Health Statistics
www.bls.gov/iif
202-691-6185
measure.alex@bls.gov