Autocoding the Survey of Occupational Injuries and Illnesses – 5 years in

Alexander Measure

Survey of Occupational Injuries and Illnesses

Example Narrative

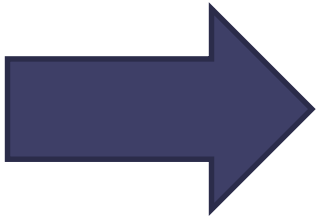
Job title: sanitation worker

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

What happened?
slipped on wet 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)

Supervised Machine Learning

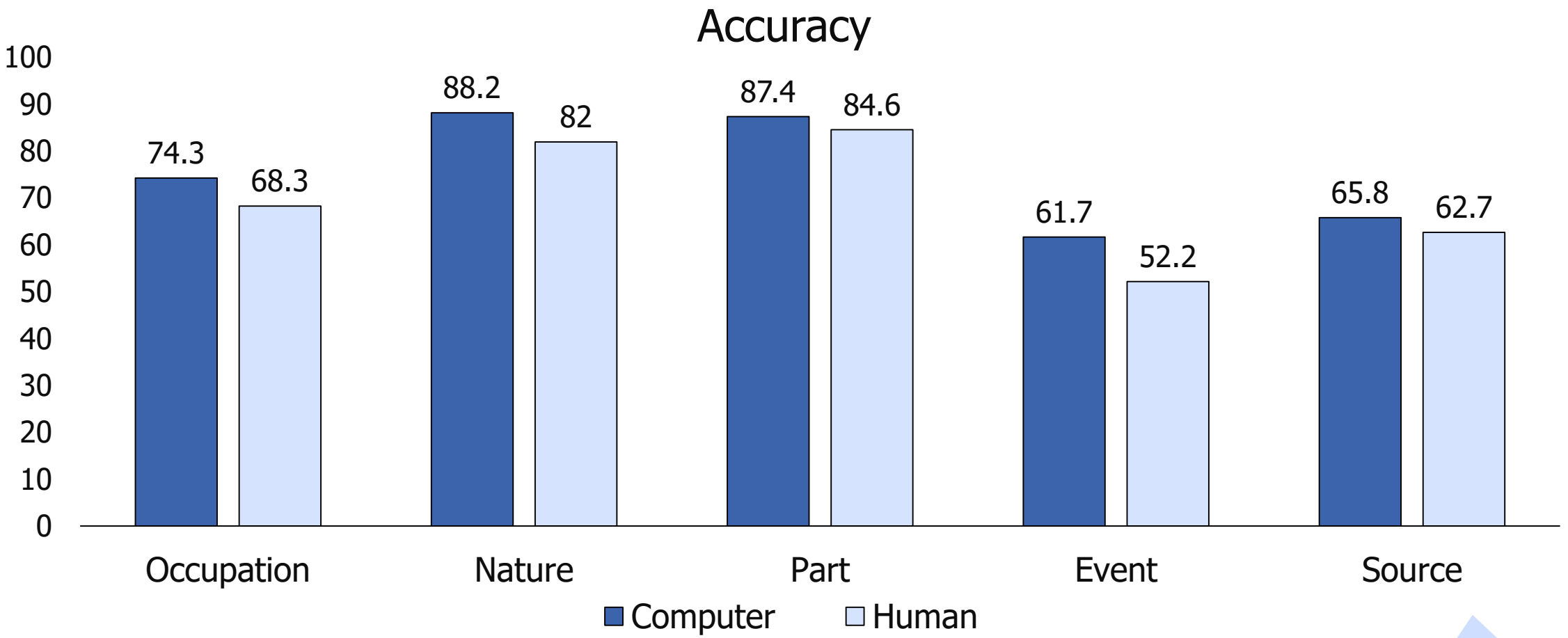
- Recipe
  - Gather previously coded data
  - Select a learning algorithm
  - Learn the autocoder from the data
- Basis for most “AI” today
  - Works well
  - Much easier to implement

```python
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.linear_model import LogisticRegression

# Read in some data
df_train = pd.read_excel('Data/msha_2010-2011.xlsx')
df_uncoded = pd.read_excel('Data/msha_2012.xlsx')

# Fit a model on df_train
vectorizer = CountVectorizer()
X_train = vectorizer.fit_transform(df_train['NARRATIVE'])
model = LogisticRegression()
model.fit(X_train, df_train['INJ_BODY_PART'])

# Autocode df_uncoded
X_uncoded = vectorizer.transform(df_uncoded['NARRATIVE'])
df_uncoded['AUTOCODE'] = model.predict(X_uncoded)
```

Does it Work?

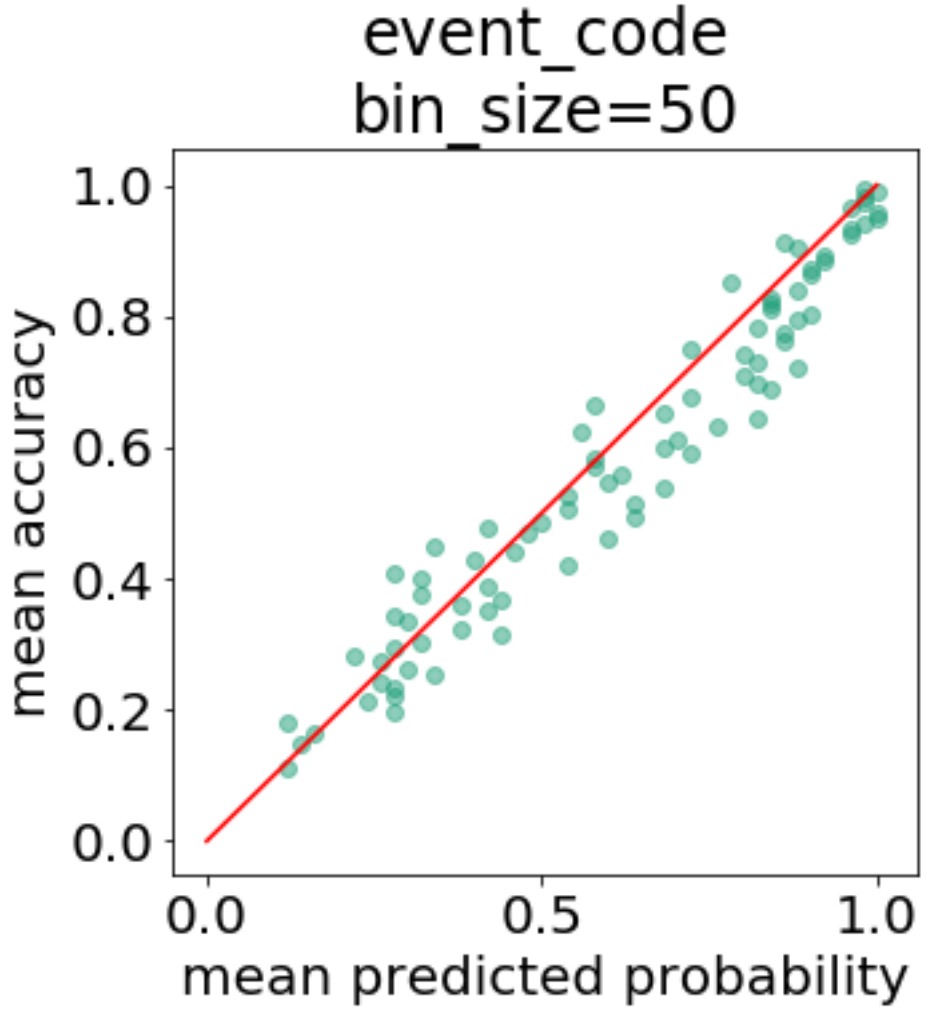
- Sample 1000 cases for “gold standard”
  - Recode each with panel of experts so we know true code
- Train autocoder on non-gold-standard data
  - Autocode gold standard
- How often does autocode match expert?
- What about manual coding process?
  - Human + regional reviewer + national reviewer + rule based edits?

Human vs. Computer Coding



The benefits of probabilistic models

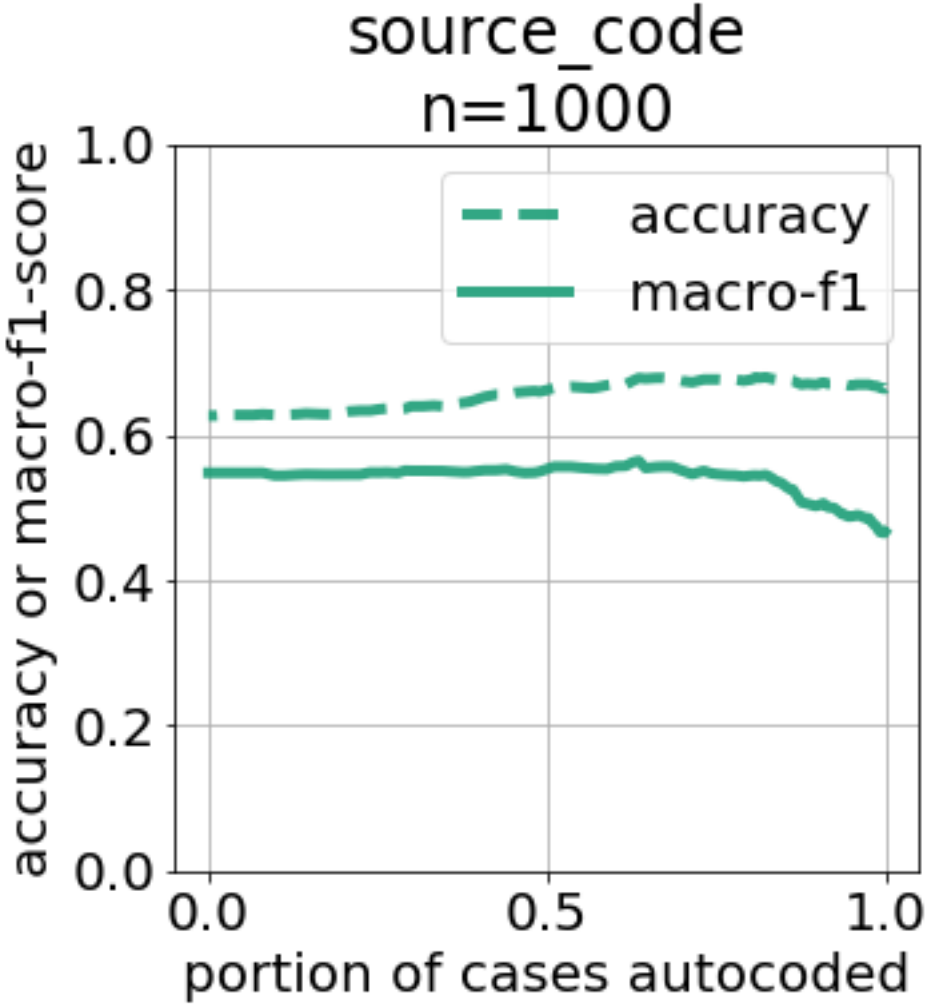
- Predicted Prob ≈ True Prob
  - It mostly knows what it doesn't know
  - Maybe a human knows?



Finding the right threshold

- For each threshold between 0 and 100%
  - If probability is above threshold, use autocode
  - Otherwise use human code
  - Evaluate resulting codes against gold standard
  - Repeat

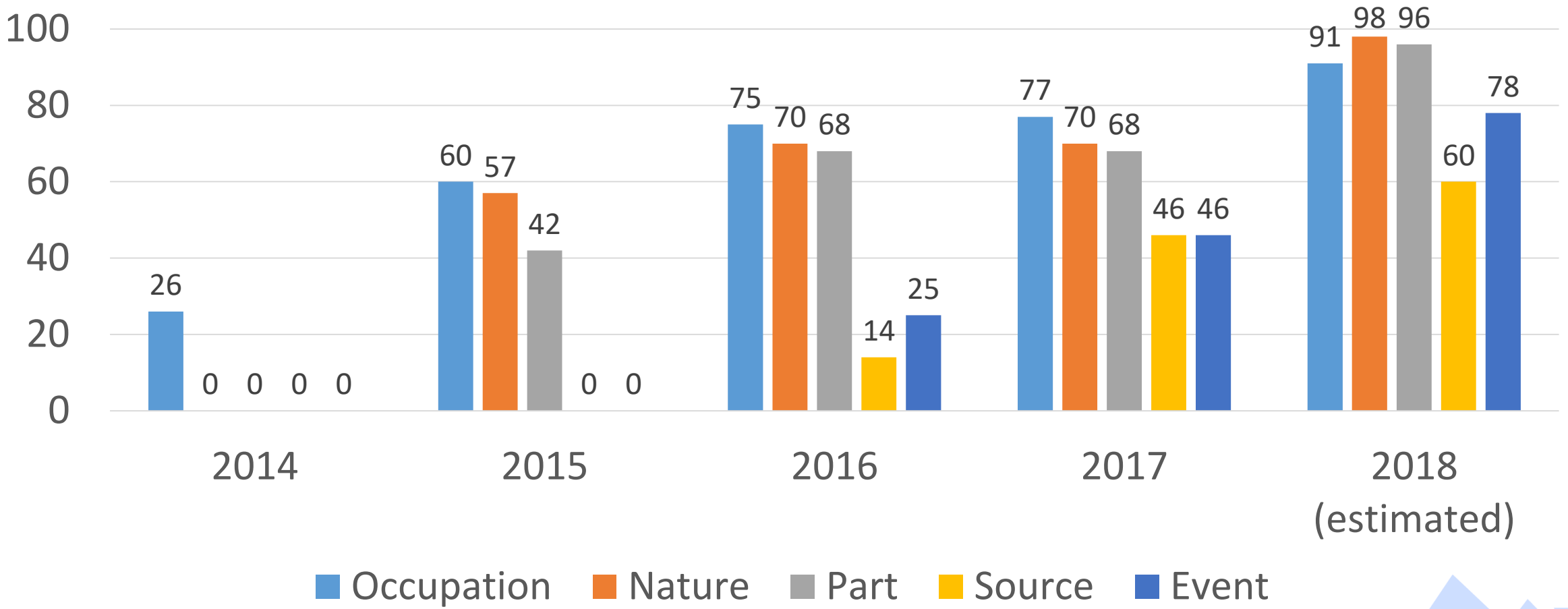
- Which threshold produces best overall quality?



What if something unexpected happens?

- Move slowly
- Keep humans in the loop
- Hold back a sample of cases and continually reassess

% of codes automatically assigned to SOII



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