Increasing survey response rates and decreasing costs by combining numeric and text mining strategies on survey paradata

Sudip Bhattacharjee (Presenter), Sudip.Bhattacharjee@uconn.edu
Senior Research Fellow, US Census Bureau; Professor, University of Connecticut, USA

Nevada Basdeo, US Census Bureau
Ugochukwu Etudo, University of Connecticut, USA; US Census Bureau
Sara Alaoui, Gunnison Consulting Group

United States Census Bureau | U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census.gov

1

Research Problem from American Community Survey (ACS) Operations

- Multi-dimensional problem
  - Declining response rates
  - Increasing collection costs
  - Exceeding respondent burden
- Multi-objective optimization problem with conflicting objectives
  - We present first steps to solve problem

United States Census Bureau | U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census.gov

2

2

Physical Contact Attempts for Final Outcome

- Outcome code 201: Occupied } Completed
- Outcome code 218: Respondent refusal
- Outcome code 313: Respondent burden exceeded } Non-interview

- Average contact attempts:
  - 201: approx. 2.5 } 1.5M contact attempts (2017 + 2018)
  - 218: approx. 5
  - 313: approx. 7 } 600,000+ contact attempts (2017 + 2018)

United States Census Bureau | U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census.gov

3

3

Costly contact attempts: Can we do better?

- **Research Question:**
  - Can we identify non-respondents based on first contact ONLY?
- **Answer:**
  - **We can identify 70-80% of non-respondent households**
- **Impact:**
  - Prioritize cases with higher probability of completion
  - Create adaptive design rules based on model results
- **How do we do it?**

United States Census Bureau | U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census.gov

4

4

Current use of CHI in ACS

- CHI – Contact History Information
  - Paradata recorded by field rep when contact attempt is not successful in getting a response
- Burden score calculation based on CHI
  - Updated based on each contact attempt
- Burden score or CHI are not used to predict final response propensity (completion or refusal)

Research question and solution approach

- **Predict respondent refusal from first contact only -- using both numeric and textual information (structured and unstructured)**
  - CHI: Structured numeric information
  - Case Notes: free form text, unstructured
  - Combine and use CHI and Case Notes
- CHI based response propensity prediction model (new for ACS)
- Case Notes based response propensity model (new for Census)

Data merging: CHI with Case Notes

- 2017 and 2018 ACS CHI and Case Notes (focus of current analysis)
- Each CHI record was merged with zero-to-many Case Notes associated with that contact attempt
- Challenge: CHI and Case Notes captured on different systems
  - Merged on control number, date and timestamp
  - Timestamp does not match
  - Manually verified large (>400) samples to identify pattern of linkage
  - Custom linkage algorithm based on control number, date and proximity of timestamp between CHI and Case Notes

Data merging with Workload table

- First contact only model:
  - **CHI + Case Notes (First contact only) → ControlNumber → FINAL OUTCOME from Workload table**
  - (completed/refused when it happens in second or later contact)
- **Focus: Predict final outcome → based on first contact info**

Distribution of Outcome Codes for First Contact only model (2017 and 2018 data)

Outcome code	Definition	2017	2017 (percent)	2018	2018 (percent)
201	Occupied	225000	45%	205000	43%
218	Respondent refusal	36500	7%	45500	10%
313	Respondent burden exceeded	13000	3%	11000	2%
301	Vacant	95000	19%	88000	18%
501	Temporary occupied	2500	0.5%	2400	0.5%
203	Sufficient partial (occupied) - no follow-up	11500	2%	11500	2%

Predictive Model Setup

- **Predict:** Final outcome
  - 201 (completed) vs 218 (refused)
  - 201 (completed) vs 313 (burden exceeded)
- Prediction based on information from 1st contact only (personal or telephone)
- **Predictors:**
  - **Model 1:** CHI only
  - **Model 2:** Case Notes only (textual data)
  - **Model 3:** ALL CHI and Case Notes
  - **Model 4 variations**: Different CHI and Case Notes variables (based on variable importance)
    - Dimensionality reduction based on chi-square selection: all features (~10,000), best 6000 features, best 4000, best 2000.

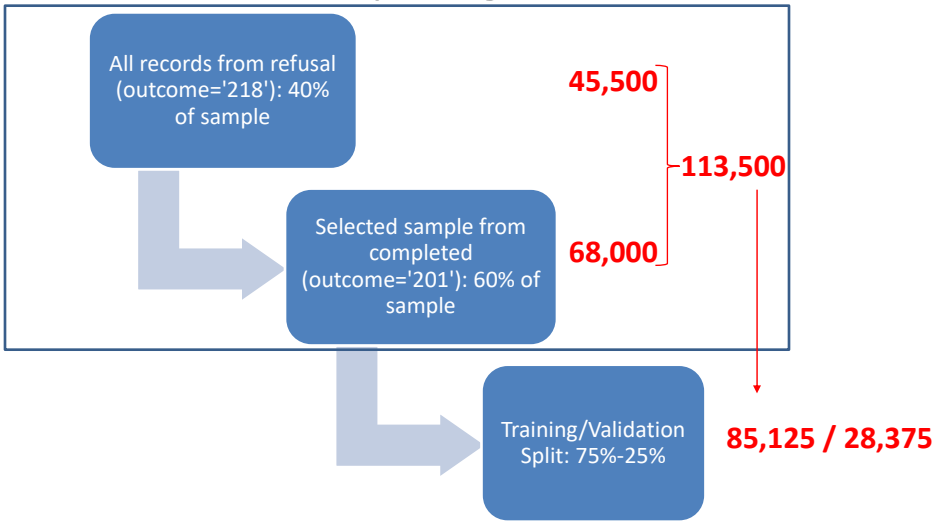
Data undersampling used for modeling

- Models run: 201 vs 218, and 201 vs 313
- "Rare" occurrence of 218 (≤ 10%) and 313 (< 5%)
  - Undersampling needed for data for modeling

Undersampled data used for modeling						
	2017			2018		
Undersampling ratio:	**50-50**	**40-60**	**30-70**	**50-50**	**40-60**	**30-70**
Outcome codes						
218	36,500	36,500	36,500	45,500	45,500	45,500
201	36,500	55,000	85,500	45,500	68,000	106,000
313	13,000	13,000	13,000	11,000	11,000	11,000
201	13,000	20,000	31,000	11,000	16,500	26,000

11

Prediction model setup (e.g. for 40/60 in 2018)



12

Methods

- NLP (natural language processing)
  - TF-IDF vectorization (Term Freq, Inverse Document Freq)
- Machine Learning Models used
  - Logistic Regression (LR)
  - Random Forest (RF)
  - Gradient boosting – XG Boost (XGB)
  - Neural Network – Multi Layer Perceptron (MLP)
  - Support Vector Machine (SVM)

- **Accounts for procedural – or model – bias in results**

Model metrics

- **Accuracy:** number corrected predicted / total (n)
- **Precision:** true refusals / total refusals predicted
  - How many predicted refusals are actual refusals?
  - Good metric, if cost of wrong prediction of refusals is high
- **Recall:** true refusals predicted / total actual refusals
  - How many of the actual refusals have been predicted?
  - Good metric, if cost of gathering survey response is high
- **F1:** $F1 = 2 \times \frac{Precision * Recall}{Precision + Recall}$
  - Measures balance between precision and recall

Best model for respondent burden (201 vs. 313)

2017

	Random Forest											
	50-50				60-40				70-30			
Features used	**Accuracy (A)**	**Precision (P)**	**Recall (R)**	**F1-value**	A	P	R	F	A	P	R	F
CHI only	0.701	0.785	0.548	0.645	0.715	0.705	0.496	0.582	0.770	0.707	0.393	0.505
Notes only	0.739	0.811	0.618	0.702	0.812	0.804	0.701	0.749	0.852	0.880	0.584	0.702
CHI + Notes (All Features)	0.790	0.765	0.856	0.808	0.832	0.808	0.761	0.784	0.864	0.833	0.683	0.751
CHI + Notes (best 2k)	0.798	0.772	0.862	0.814	0.829	0.799	0.767	0.783	0.864	0.827	0.689	0.752

2018

	50-50				60-40				70-30			
Features used	**Accuracy (A)**	**Precision (P)**	**Recall (R)**	**F1-value**	A	P	R	F	A	P	R	F
CHI only	0.721	0.682	0.860	0.761	0.715	0.705	0.496	0.582	0.770	0.707	0.393	0.505
Notes only	0.757	0.742	0.809	0.774	0.812	0.804	0.701	0.749	0.852	0.880	0.584	0.702
CHI + Notes (All Features)	0.790	0.765	0.856	0.808	0.832	0.808	0.761	0.784	0.864	0.833	0.683	0.751
CHI + Notes (best 2k)	0.798	0.772	0.862	0.814	0.829	0.799	0.767	0.783	0.864	0.827	0.689	0.752

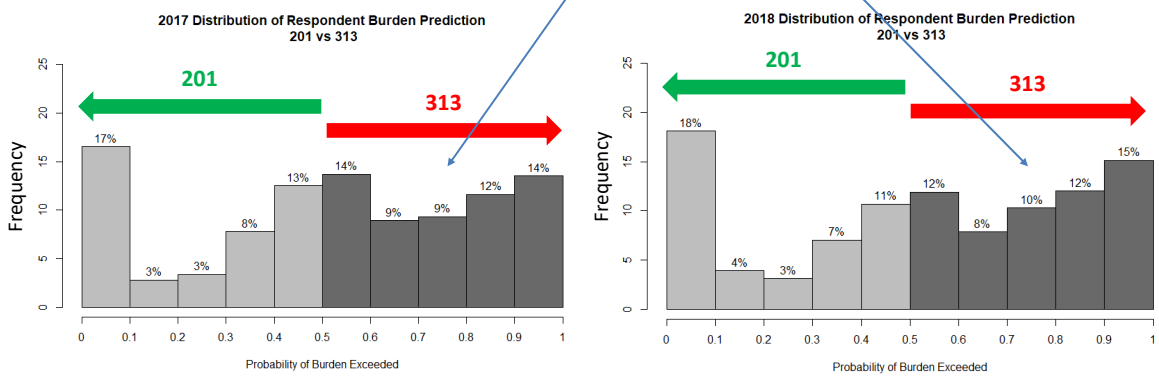
United States Census Bureau | U.S. Department of Commerce, Economics and Statistics Administration, U.S. CENSUS BUREAU, census.gov

15

15

Distribution of Probability of Final outcome based on first contact only (201 vs. 313)

Differentiated strategies can be implemented for these segments



Model with CHI + Notes (All features), 50-50 undersampling ratio

United States Census Bureau | U.S. Department of Commerce, Economics and Statistics Administration, U.S. CENSUS BUREAU, census.gov

16

16

Best model for respondent refusal (201 vs. 218)

- Similar as in 201 vs. 313 modeling
- Best model:
  - CHI + Notes (all features), Random Forest, 40-60 undersampling
- Notes-only model accuracy 20-25% better than CHI-only model
- Probability distribution of final outcome can provide differentiated strategies for operational implementation

Feature importance

- Mix of CHI and Case Notes terms (single and double-word phrases)
- Can be used to spot refusal reason trends in different times and geographies
- Can be used to train FR

Transfer learning
Response propensity – 201 vs. 218

- Train on 2017 data, predict 2018
- Choose best trained model from 2017
  - Random Forest, 50-50 undersampling
- Predict refusals (218) for 2018
  - 2018 Jan-Mar
  - 2018 Jan-June
  - 2018 full year
- Insights:
  - Model accuracies dropped, as expected
  - Need to build rolling horizon model
  - Add state and RO (geographic dimension)
  - Add month of survey (time dimension)

Deep learning modeling ongoing

- Predictors:
  - CHI
  - Case Notes
  - State and RO (regional office) } geographic dimension
  - Survey month } time dimension
- Methods:
  - NLP: NER (Named Entity Recognition), ELMo, BERT

Conclusion and next steps

- Promising results from CHI + Case Notes predictive model
  - Used only 1st contact information to predict eventual outcome
  - Augmenting with newer datasets (2019, 2020)
- Can provide reliable recommendations for eventual refusal cases
- Provide highly confident refusal recommendations
  - Lower the data collection priority on predicted cases (eventual refusals)
  - Add a high value to burden score
- Can lead to savings in data collection
  - Cost of each contact/case
- Continue to fine-tune models
  - Deep learning models

- Experiment: For medium confident refusal recommendations, use some treatment to see if it increases response rate