Automated Coding of Political Campaign Advertisement Videos: An Empirical Validation Study

June Hwang[†] Kosuke Imai[†] Alex Tarr[‡]

[†]Harvard University [‡]Princeton University

Applied Statistics Workshop

Harvard University

April 3, 2019

Motivation

- Modern political campaigns rely on various kinds of advertisements
- In 2018, TV ads were the most popular medium $\rightsquigarrow$ \$8.5 billion
- Questions:
  1. How do campaigns choose the contents of ads?
  2. How do the contents of ads affect the behavior and opinion of voters?

- Main data source on TV ads: Wesleyan Media Project (WMP)
  - successor to the Wisconsin Advertisement Project (WAP)
  - all federal and gubernatorial elections from 1998 to 2016
  - videos obtained from the Campaign Media Analysis Group (CMAG)
  - a group of research assistants code over 100 variables:
    1. CMAG: broadcast time and frequency, media market, TV show, etc.
    2. WMP: issue mentions, opponent appearance, negativity, etc.
- Data not publicly available until the next election

Overview of the Project

- Goals:
  1. Automate the coding of campaign advertisement videos
  2. Compare the results of automated coding with those of human coding
- Workflow:
  1. Data acquisition $\rightsquigarrow$ audio matching
  2. Feature construction
     - visual features: video summarization, image text detection, face detection
     - audio features: speech transcription, text features, music features
  3. Empirical validation
     - issue mention, opponent mention, face recognition
     - music mood classification, negative advertisement
- Findings:
  1. Machine coding is at least as accurate as human coding
  2. In some cases, machine coding is too accurate
  3. Music mood and negativity classifications have a room for improvement

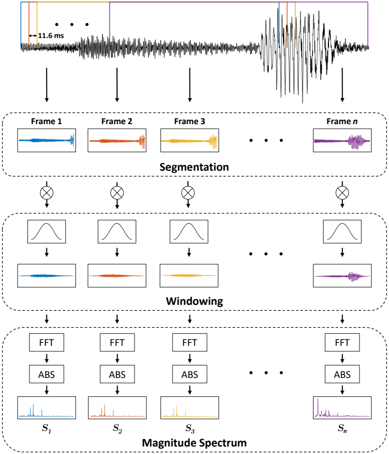
Data Acquisition from YouTube

- CMAG videos have low resolution images and low quality audio $\rightsquigarrow$ unsuited for machine coding
- High resolution videos from candidates' official YouTube channels
- Filter by length (15, 30, and 60 seconds $\pm$5 seconds)

Election cycle	Office	All candidates	Candidates with YouTube channels		All videos
2012	President	2	2	(100%)	400
	House	317	263	(83.0%)	1225
	Senate	64	50	(78.1%)	683
	Governor	25	20	(80.0%)	194
2014	House	255	199	(78.0%)	1047
	Senate	68	52	(76.5%)	997
	Governor	86	59	(68.6%)	888
	Total	817	645	(79.0%)	5434

Matching YouTube Videos with CMAG Videos

- Direct comparison of automated coding with the WMP coding requires matching of YouTube videos with CMAG videos
- Audio matching based on spectrogram (Haitsma and Kalker 2002)
  1. split audio signal into 31/32 overlapping segments $\rightsquigarrow$ 11.6ms per segment
  2. windowing to reduce noise due to segmentation
  3. Fast Fourier transform (FFT)
  4. Absolute value transform (ABS)
- Dimension reduction via energy values $\rightsquigarrow$ spectral fingerprint
- Matching on sub-fingerprint
- Evaluation: a random sample of 50 matches and 50 non-matches



The Validation Data Set

		All Candidates		Republicans		Democrats	
Election cycle	Office	CMAG videos	Matches found	CMAG videos	Matches found	CMAG videos	Matches found
2012	President	228	80.7%	98	71.4%	130	87.7%
	House	1106	54.7	574	49.7	506	63.0
	Senate	586	55.0	279	45.5	289	65.1
	Governor	184	54.4	94	48.9	90	60.0
2014	House	912	57.7%	437	57.7%	470	58.3%
	Senate	666	71.3	327	70.3	307	76.5
	Governor	742	51.6	383	49.1	317	59.3
	Total	4424	58.7%	2192	54.7%	2109	65.1%

- better coverage for presidential candidates, Democrats, 2014 elections
- regression analysis $\rightsquigarrow$ incumbency (channel), partisanship (videos)

Video Summarization

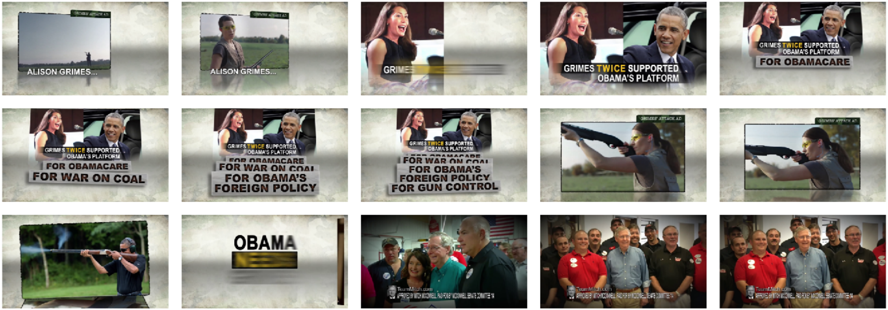
- Video data = a sequence of *frames*
- YouTube data have 24 or 30 frames with 1280×720 pixels per second $\rightsquigarrow$ a total of 720 – 1,800 frames (or several gigabytes) per video
- Need to select a small number of representative frames
- Video summarization algorithm (Chakraborty *et al.* 2015)

$$S^* = \underset{S \subseteq V}{\operatorname{argmax}} \underbrace{\sum_{i \in V} \max_{j \in S} w_{ij}}_{\text{representativeness}} + \lambda_1 \underbrace{\sum_{i \in S} \min_{j \in S} d_{ij}}_{\text{uniqueness}} + \lambda_2 \underbrace{(N - N_S)}_{\substack{\#\text{ of unselected} \\ \text{frames}}},$$

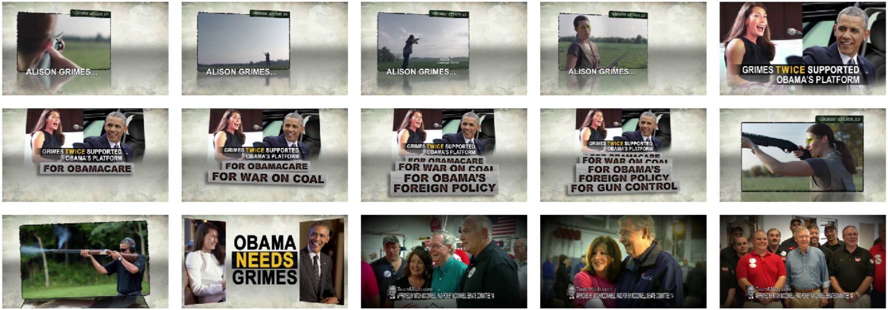
  - V: frames of original video data
  - S: set of selected frames
  - w_{ij}: cosine similarity of histogram of oriented gradients (HOG)
  - d_{ij}: χ^2 distance based on the Lab histogram
- Approximate optimization algorithm

Ad for Mitch McConnell (Rep. Sen. KS; 2014)

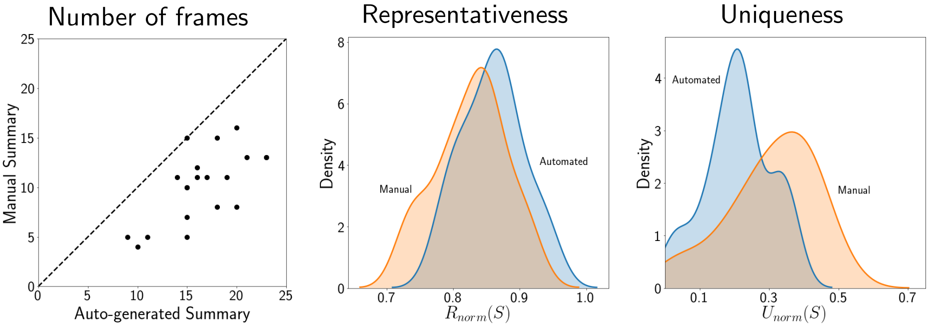
auto-generated summary



manually-generated summary



Auto-generated vs Manually-generated Summaries



- More frames for auto-generated summaries
  ⇝ more representative but less unique

Image Text Detection

- Google Cloud Platform (GCP) Vision API

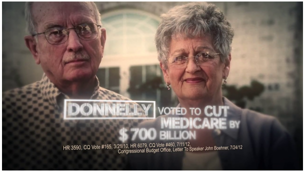
(a) Newspaper

(b) Background image

(c) Endorsement

(d) Approval message

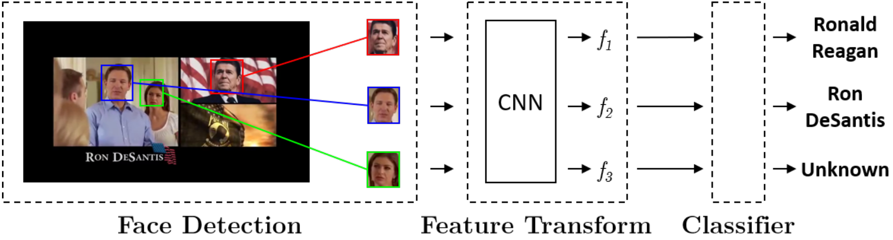
(e) Voting records



(f) Policy position

- (a), (b), (c) $\rightsquigarrow$ perfect detection
- (d), (e), (f) $\rightsquigarrow$ missing a few words

Face Detection



- Multi-task cascade neural networks (MTCNN; Python package facenet) with the loss function (Zhang *et al.* 2016):

$$\sum_{i=1}^{N} -\left\{d_i \log \hat{d}_i + (1-d_i)(1-\log \hat{d}_i)\right\} + \frac{\mathbf{1}\{d_i=1\}}{2}\left(\|b_i-\hat{b}_i\|^2 + \|l_i-\hat{l}_i\|^2\right)$$

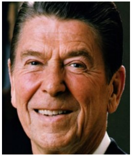
  - d_i: binary variable indicating the presence of face
  - b_i: bounding box for face
  - l_i: facial landmark locations
  - "hat" represents predicted value from the MTCNN
  - WIDER FACE and CelebA data sets as training data

Facial Features

- FaceNet algorithm (Schroff *et al.* 2015)
  - convolutional neural nets
  - uses Google's Inception ResNet V1 architecture
  - trained on the VGGFace2 data set (several million face images)
- Triplet loss function to learn about embedding $f(x_i) \in \mathbb{R}^{128}$:

$$\sum_{j=1}^{N_{\text{trip}}} \max\left(0, ||f(x_j^a) - f(x_j^p)||^2 - ||f(x_j^a) - f(x_j^n)||^2 + \alpha\right)$$

  - x_j^a: anchor image
  - x_j^p: positive image, i.e., the same person as x_j^a
  - x_j^n: negative image, i.e., different person
- Hard-to-classify triplets:



Anchor image (x_j^a)

Positive image (x_j^p)

Negative image (x_j^n)

Speech Transcription

- Google Cloud Platform Video Intelligence API
  - Recurrent neural network called Long short-term memory (LSTM)
  - Known to be accurate (Prabhavalkar *et al.* 2017)
  - Political science validation (Proksh *et al.* 2019)

- Works well for ads too:

Auto transcription	Manual transcription
"...it's about getting new jobs getting good jobs given middle class people the chance to get her kids a decent life nobody can tell me it's not a senator's job to create jobs and I choose Allison because she will work with people in both parties to do what's right for you since Alison to the Senate"	"...it's about getting new jobs getting good jobs giving middle class people the chance to give their kids a decent life nobody can tell me it's not a senator's job to create jobs and I choose Alison because she will work with people in both parties to do what's right for you send Alison to the Senate"

- A small number of mistakes: songs, kids' voice, etc.

Ad for Joe Dorman (Dem. Gov. OK; 2014)

- **Transcript:**

  I'm not fir gun control yes I'm f***ing control but I'm becoming car no I'm not common what did I say before I don't know anymore nobody's keeping score

- **Image text:**

  I'M NOT FOR GUN CONTROL
  YES, I'M FOR GUN CONTROL
  MMON ORE BUT I'M OR COMMON CORE
  NO, I'M NOT FOR COMMON CORE
  WHAT DID I SAY BEFORE?
  I DON'T KNOW, ANYMORE I DON'T KNOW ANYMORE
  HOPE NOBODY'S KEEPING SCORE
  CONSISTENCY IS SUCH A BORE
  FLIP-FLOP FALLIN
  PAID FOR BY JOE DORMAN FOR GOVERNOR

Ad for Nan Hayworth (Rep. House. NY18; 2014)

- **Transcript:**

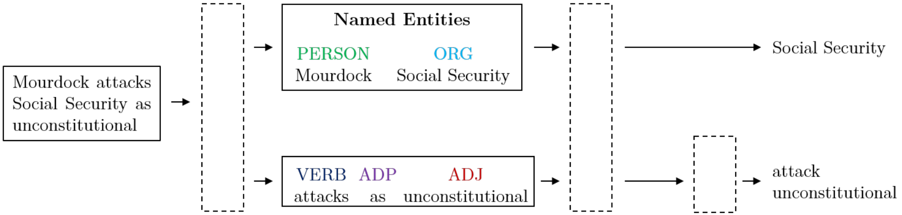
  sean malone is a phony baloney baloney baloney is full of baloney that's right shaun maloney is a phony baloney baloney making big promises but then voting to cut medicare and veterans pensions a phony baloney pony big big phony and while we struggle maloney voted for amnesty for illegals amnesty amnesty really and first class airfare for congress said is right definitely shaun maloney is full of baloney baloney baloney head in washington i'm nan hayworth and i approved this message

- **Image text:**

  SEAN Maloney Phoi
  ECI THF US FOUND TO BE UNTRUTHFUL
  one
  SEAN Maloney CUT Medicare
  CUT Medicare CUT Veteran's Pensions
  SERN Malonev Amnesty for Illegals
  E Maloney FREE First Class Airfare
  672
  NAN CONGRESS PAID FOR BY FRIENDS OF NAN HAYWORTH.APPROVED BY NAN HAYWORTH DOCTOR. MOTHER. NEIGHBOR

Text Features

- Keyword based approach $\rightsquigarrow$ issue and opponent mentions
- Machine learning for sentiment analysis $\rightsquigarrow$ negativity
- Pre-processing transcripts (Python package spacy):
  - part-of-speech tagging and named entity recognition using LSTM (Dozat and Manning 2016)
  - lemmatization rather than stemming
    - "caring" $\rightsquigarrow$ "care" instead of "car"
    - recognizes "mice" as a plural of "mouse"



Music Features

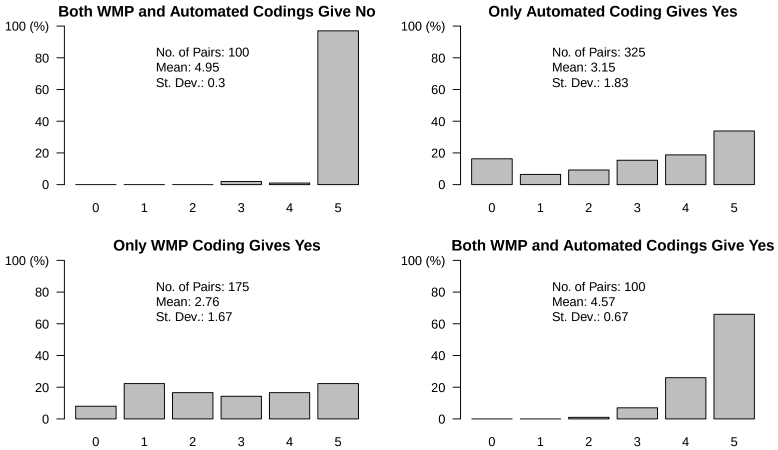
- Music is important for tone of an advertisement
- WMP's variable for music mood:
  1. ominous and tense
  2. uplifting
  3. sad and sorrowful
- Use of spectrogram as done for audio matching (Ren *et al.* 2015)
- We do not separate music and speech but compute features that are known to characterize types of music well
- 412 short-term features:
  1. Statistical spectrum descriptor (SSD): shapes of spectrogram
  2. Mel-frequency cepstral coefficients (MFCC): energies
  3. Octave spectral contrast (OSC): differences in the peaks and valleys
  4. Spectral flatness measure and spectral crest measure (SFM/SCM)
- 224 long-term features:
  1. Modulation feature spectrogram: rhythm, tempo, and beat
  2. Joint-frequency feature: temporal evolution of modulation features

Issue Mention

- Whether an ad mentions or pictures certain political issues or actors
- A key set of variables in the WAP/WMP data sets
  1. 10 actors: Obama, Pelosi, McConnell, Democrats, Republicans, ...
  2. 12 politically-charged words: tea party, wall street, big government, ...
  3. 61 issues: tax, jobs/employment, gun control, drugs, ...

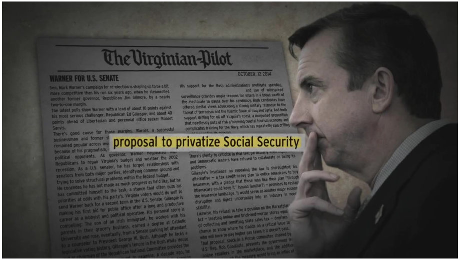
- keyword based search
  - 44 issues: we use the WMP issue names and last names of actors
  - 16 issues: we add synonyms and words with the same roots (e.g., "Chinese" for the "China" issue, "farm" for the "farming" issue)
  - 21 issues: we add relevant words (e.g., "climate change" for "global warming", "NRA" for the "gun control" issue)
- No stemming and no lemmatization

		Automated coding			
		Audio Data Only		Audio and Visual Data	
		No	Yes	No	Yes
WMP coding	No	197,986 (95.72%)	1,501 (0.73%)	197,173 (95.33%)	2,314 (1.12%)
	Yes	1,776 (0.86%)	5,573 (2.69%)	1,488 (0.72%)	5,861 (2.83%)



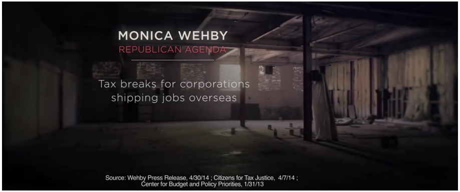
Examples of Mistakes by Automated Coding

Ad for Mark Warner (Dem. Sen. VA; 2014)



- Reading the entire excerpt from the newspaper
- Incorrectly choosing the "tax" issue

Jeff Merkley (Dem. Sen. OR; 2014)



- Detected the word "budget" from the name of the organization quoted as the source

Opponent Mention

- The WMP excludes the oral approval: "Excluding the *oral approval*, is the opposing candidate mentioned by name in the ad?"
- We use last name (Roe), possessive (Roe's), and possessive without an apostrophe (Roes)
- Results:

		Automated coding			
		Audio Data Only		Audio and Visual Data	
		No	Yes	No	Yes
WMP coding	No	1,273 (51.43%)	64 (2.59%)	1,260 (50.91%)	77 (3.11%)
	Yes	127 (5.13%)	1,011 (40.85%)	28 (1.13%)	1,110 (44.85%)

1. 77 "false positives": 3 mistakes by automated coding (detecting texts in background image)
2. 28 "false negatives": 18 mistakes by automated coding (mistakes in transcription or image text detection)

Face Recognition

- We combine two WMP variables:
  1. "Excluding the *oral approval*, is the favored candidate / opposing candidate pictured in the ad?"
  2. "Does the candidate physically appear on screen and speak to the audience during oral approval?"
- This is supposed to exclude the case where the candidate appears but does not speak $\rightsquigarrow$ we do not make this distinction
- 75 Senate candidates from 2012 and 2014 elections
- Scraped images from Wikipedia and other pages on the Internet

		Automated coding			
		Favored candidate		Opposing candidate	
		No	Yes	No	Yes
WMP coding	No	58 (7.56%)	109 (14.21%)	490 (63.89%)	12 (1.56%)
	Yes	57 (7.43%)	543 (70.80%)	65 (8.47%)	200 (26.08%)

1. 166 disagreements for the favored candidate
   - 94 cases: detected in the oral approval segments
   - 48 cases: angled, occluded, and dimly-lit images
   - 24 cases: mislabels by the WMP coders
2. 77 disagreements for the opposing candidate
   - 51 cases: angled, occluded, and dimly-lit images
   - 26 cases: mislabels by the WMP coders

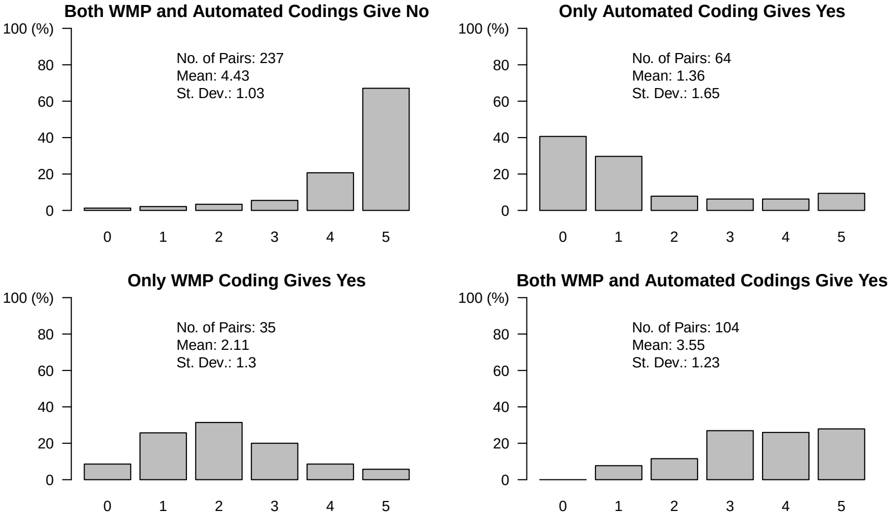
Music Mood Classification

- Original WMP question: "If music is played during the ad, how would it be best described?"
- Out of 2,276 videos,
  - "uplifting" 70%, "ominous/tense" 32%, "sad/sorrowful" (14%)
  - 15% have more than one category
- SVM classifier with radial basis and 5-fold cross validation

		Automated coding					
		Ominous/Tense		Uplifting		Sad/Sorrowful	
		No	Yes	No	Yes	No	Yes
WMP	No	237 (53.86%)	64 (14.55%)	66 (15.00%)	65 (14.77%)	334 (75.91%)	45 (10.23%)
	Yes	35 (7.95%)	104 (23.64%)	31 (7.05%)	278 (63.18%)	31 (7.05%)	30 (6.82%)

- WMP intercoder (2 coders) agreement rate: 84 – 92%
- state-of-the-art machine learning methods $\rightsquigarrow$ 70% accuracy

MTurk Study for the "Ominous/tense" Question



- 85% agreement rate between the WMP coding and the majority opinion of MTurkers

Negativity

- CMAG variable: "positive," "negative," and "contrast"
- WMP's original question: "In your judgment, is the primary purpose of the ad to promote a specific candidate, attack a candidate, or contrast the candidates?" — "contrast", "negative", and "attack"
- We focus on "positive" vs. "negative" from the CMAG
- Liner SVM with 3-fold cross validation

		Automated coding					
		Text Only		Music Only		Text and Music	
		Negative	Positive	Negative	Positive	Negative	Positive
WMP	Negative	291 (56.18%)	34 (6.56%)	255 (49.23%)	70 (13.51%)	290 (55.98%)	35 (6.76%)
	Positive	43 (8.30%)	150 (28.96%)	63 (12.16%)	130 (25.10%)	39 (7.53%)	154 (29.73%)

- Need to tune music features for dark music

Concluding Remarks

- Many variables form the WAP and WMP can be automatically coded
  - Often, machine coding is as accurate as human coding
  - Music mood and negativity classifications have a room for improvement
- We can improve the efficiency and scope of research on political advertising (TV, radio, and online)
- Video data = audio data + image data + text data
- WAP and WMP serve as excellent validation data sets
- Contribute to the fast growing political science literature on analyses of audio, image, and transcript data (e.g., Dietrich, 2018; Dietrich *et al.* 2018; Knox and Lucas, 2018; Proksch *et al.* 2019; Torres, 2018)
- Our code will be made available

Send comments and suggestions to
Imai@Harvard.Edu