DEMOCRATIC DEBATE VIEWPOINT EXTRATOR

Information Retrieval and Web Agents, Spring 2016

Tom Catullo & Vamsi Chunduru

Overview

The Democratic Debate Viewpoint Extractor was created in congruence with the 2016 Democratic Primary season. As the Democratic National Convention on July 25 rapidly approaches, there are still two candidates very much in the running, Bernie Sanders and Hillary Clinton. The purpose of the DDVE is to identify each candidate's stance on popular political topics, especially as time has passed throughout the primary season.

The DDVE allows the user to select between the two runners, and query based on political topics to extract statements uttered by either candidate during any of the 2016 Democratic Debates. The tool applies vector-IR concepts to identify relevance of statements said during the debates, and reports the results in order of importance.

Our goal is to provide an easy interface that users can utilize to become more knowledgeable about our Democratic candidates, and in turn become more involved in the political process.

Methodology

Data Collection & Preparation

Our initial approach was to simply collect texts of speeches made by each candidate and analyze those. However, we ran into difficulty in 1) acquiring these texts and 2) consistently processing them. Furthermore, it was less likely that the speeches would line up in such a way that would provide insightful comparisons between candidates. Our solution was to use debate transcripts, which were easily accessible from *The American Presidency Project, University of California, Santa Barbara*¹. Another benefit of this data set was that all the transcripts were formatted in the same manner, so the process of mining through each transcript to extract statements was streamlined².

Once we had the files, the final step was to segment the texts based on the candidates and split them into separate files. At this stage our raw data was ready to be processed.

Data Processing

The next step in our project was to create a vector-IR system given the pre-processed dataset. We treated each section of text (usually an individual statement made by a candidate) as a document. The intuition behind this was that the context of a topic within a statement would remain the same, however

¹ http://www.presidency.ucsb.edu/debates.php

² Refer to file *splitter.py* in Appendix for details

that can change between statements. With the segmented documents, we constructed document vectors using raw term frequency weighting.

Topic Vectors

To create topic vectors, we initially read through a couple debate transcripts and identified prevalent topics. For each topic, we created a list of pertinent keywords:

T: infrastructure	T: Wall Street	T: gun violence	T: immigration
bridges	big banks	assault weapon	citizenship
roads	corporate greed	NRA	border
	bailout	gun control	visa
			H-1B
			H1B

Our approach was to create a preliminary list and then expand the set of topics and keywords after extracting results. Below is the updated list:

T: infrastructure	T: college affordability	T: healthcare	T: corporate greed	T: wealth inequality	T: gun control	T: immigration
bridges	college	health	banks	wealth	NRA	citizenship
roads	colleges	care	corporate	inequality	gun	border
water	education	insurance	greed	middle	guns	visa
pollution	public	universal	bailout	class	control	H-1B
crumbling	tuition	Medicare	Wall	millionaire	violence	H1B
services	university	prescription	Street	millionaires	murder	refugee
poisioning	universities	premiums		billionaire	kill	undocumented
poisioned	student	drug		billionaires	killing	deportation
resources	debt			unemployment	weapon	deported
teachers	debt-free			wage	weapons	guest-worker
	degree			wages	tragedy	guestworker
				income	assault	guestworkers
				rigged	gunmakers	
					ammunition	

We also implemented dynamic topic vector creation using user input regarding topic keywords. This uses the assumption that the user is knowledgeable about the topic and can provide a relevant set of keywords.

With both vector sets in place, we can perform cosine similarity measures between each document vector and the given topic vector. This is computed on the spot, when a search of a given topic and candidate is run by the user.

Results & Analysis

As specified above, we focused on two democratic candidates and seven topics (starting at four topics). We found that as we expanded the set of keywords, the accuracy of the relevance ranking

improved considerably, producing more appropriate results for each topic. Although this may seem selfevident, it was a good validation that our approach was effective and achieving our goal.

For example, we can compare the results from the topic of "Infrastructure." With its limited list of keywords, we encountered some anomalous results:

```
EXCERPT #5:
-------
Wolf, thank you very much. CNN, thank you very much. Secretary Clinton, thank you very much.
```

But once we expanded the keywords, the same result was much more accurate:

Refer to the Appendix for more detailed outputs.

Conclusion

All in all, we found that even fundamental vector-IR algorithms, i.e. cosine similarity and raw term weighting, are considerably effective at identify contexts within large corpus data, and retrieving query-relevant information. This project lays the foundation for a broader scope of application for making dense text data more accessible to the average user. Especially within the context of political discourse, a barrier to entry for many newcomers has been the density of political text that requires large amounts of mining to extract the relevant bits of knowledge. This results in laypersons to depend on biased reviews of the political activity as opposed to directly interacting with the content. We hope that tools like this can help improve the political process by drawing in more individuals and improving the overall quality of political discourse in this country.

Appendix

List of files

- 1. Process.py
 - a. This program uses the preprocessed speech document files within Speech directory (segmented based on candidate) and topic_terms.txt to build document and topic vectors and compute cosine similarity based on topic queries.
 - b. Run using script: ./process
- 2. Speeches/splitter.py
 - a. This program uses the raw debate transcripts and splits them based on candidate name.
 - b. Run using script: ./splitter DEBATE_TRANSCRIPT
 - c. Example: ./splitter DemDebate1.txt creates two files, one for Clinton and one for Sanders, and places them in their respective folders.
- 3. Topics_terms.txt
 - a. List of prepared topics and their associated keywords
- 4. DemDebateX.txt
 - a. Raw debate transcript of a given 2016 primary democratic debate
- 5. CANDIDATE_NAME/DemDebateX.txt
 - a. Segmented debate transcript based on candidate

Outputs

Top 5 results for SANDERS

TOPIC: Infrastructure

EXCERPT #1:

What is going is a disgrace beyond belief. As the president of the United States, this is what I would do is if local government does not have the resources — if state government for whatever reason America shouldn't be poisoned, federal government comes in, federal government acts. [applause] What is absolute incredible to me is that water rates have soared in Flint. You are paying three times more for poison water than I am paying in Burlington, Vermont for clean water. First thing is you say, people are not paying a water bill for poison and that is retroactive. [applause] Second of all, to ease anxiety, CDC has to come in and examine every child and adult in this community in terms of the amount of lead they may have. Thirdly, the wealthiest country has to rebuild the crumbling infrastructure, our water systems. I have a bill for a trillion dollars, it creates 13 million jobs rebuilding flint, Michigan and communities all over the country.

EXCERPT #2:

Over the last several weeks I had the opportunity to meet with a number of residents of Flint at a town meeting in Flint, and I have to tell you what I heard, and what I saw literally shattered me. And, it was beyond belief that children in Flint, Michigan in the United States of America in the year 2016 are being poisoned. That is clearly not what this country should be about. As Anderson indicated, there's a lot of blame to go around, and one of the points that I have made is that I believe the Governor of this state should understand that his dereliction of duty was irresponsible. He should resign. [applause] But more importantly, what is happening in Flint to a lesser degree is happening throughout this country. In recent years we have seen a proliferation of millionaires and billionaires, and increase in income and wealth inequality, and all over this country middle class families are struggling. Cities and towns are struggling, in order to provide basic services. Among many other things, we need to rebuild our crumbling infrastructure, our water systems, our waste

water plants, our roads and our bridges. The wealthiest country in the history of the world has got to get its priorities right, take care of the people, no more tax breaks for billionaires.

EXCERPT #3:

I will make a personal promise to you that the EPA and the EPA director that I appoint will make sure that every water system in the United States of America is tested, and that the people of those communities know the quality of the water that they are drinking, and that we are gonna have a plan to rebuild water systems in this country that are unsafe for drinking.

EXCERPT #4:

Well, to put that in a context, Judy, I think we have to understand that in the last 30 years in this country there has been a massive transfer of wealth going from the hands of working families into the top one-tenth of 1% whose percentage of wealth has doubled. In other words, the very rich are getting richer, almost everybody is going getting poorer. What I believe is the United States, in fact, should join the rest of the industrialized world and quarantee healthcare to all people. Our Medicare for all single-payer proposal will save the average middle class family \$5,000 a year. I do believe that in the year 2016 we have to look in terms of public education as colleges as part of public education making public colleges and universities tuition free. I believe that when real unemployment is close to 10%, and when our infrastructure, our roads, our bridges, our water systems, Flint, Michigan comes to mind. Our waste water plants, our rail, our airports, in many places are disintegrating. Yeah, we can create 13 million jobs by rebuilding our infrastructure at a cost of a trillion dollars.

EXCERPT #5:

Well, let's begin by discussing the problem. As you've indicated, the American Society of Civil Engineers say we need trillions of dollars — trillions of dollars — just to bring up our infrastructure to deal with — water systems like Flint, just to bring them up to decent levels. We have suggested a trillion-dollar investment. Now, you know how we're gonna pay for that? I'll tell you how we're gonna pay for

that. Right now, this country is losing \$100 billion every single year because profitable corporations are stashing their profits in the Cayman Islands, Bermuda and in other tax havens...[applause]...and in some cases, not paying five cents in federal income tax. I will eliminate that outrageous loophole, and we will raise \$1 trillion. And by the way, not only do we rebuild our crumbling infrastructure — \$1 trillion over five years creates 13 million decent-paying jobs.

Top 5 results for SANDERS

TOPIC: Immigration

EXCERPT #1:

The answer is yes. We've got 11 million undocumented people in this country. I have talked to some of the young kids with tears rolling down their cheeks, are scared to death that today they may or their parents may be deported. I believe that we have got to pass comprehensive immigration reform, something that I strongly supported. I believe that we have got to move toward a path toward citizenship. I agree with President Obama who used executive orders to protect families because the Congress, the House was unable or refused to act. And in fact I would go further. What would motivate me and what would be the guiding light for me in terms of immigration reform, Judy, is to bring families together, not divide them up. And let me say this also. Somebody who is very fond of the president, agrees with him most of the time, I disagree with his recent deportation policies. And I would not support those. Bottom line is a path towards citizenship for 11 million undocumented people, if Congress doesn't do the right thing, we use the executive orders of the president. [applause]

EXCERPT #2:

I know that. But point being is that progressives did vote against that for that reason. My view right now — and always has been — is that when you have 11 million undocumented people in this country, we need comprehensive immigration reform, we need a path toward citizenship, we need to take people out of the shadows.

EXCERPT #3:

Poverty is increasing. And if wages are going down, I don't know why we need millions of people to be coming into this country as guestworkers who will work for lower wages than American workers and drive wages down even lower than they are right now. [end video clip] [crosstalk]

EXCERPT #4:

Well, when we talk about efforts to assist immigrants, Secretary Clinton prevailed upon the governor of New York, Elliot Spitzer, who wanted to do the right thing and provide driver's license to these who were undocumented, she said don't do it, and New York State still does not do it. In Vermont, by the way, I worked with officials and undocumented people in Vermont do have the ability to get driver's license. [applause] When we talk about immigration, the secretary will remember that one of the great tragedies, human tragedies of recent years is children came from Honduras where there's probably more violence than almost any place in this country, and they came into this country. And I said welcome those children into this country, Secretary Clinton said send them back. That's a difference.

EXCERPT #5:

A couple of years ago, I had the opportunity to go on a congressional delegation. And I went to one of these Turkish refugee camps right on the border of Syria. And what a sad sight that was: Men, women, children forced out of their homes. And Turkey, by the way, did a very decent thing, providing what was reasonable housing and conditions for those people. It seems to me that given our history as a nation that has been a beacon of hope for the oppressed, for the downtrodden, that I very strongly disagree with those Republican candidates who say, you know what, we've got to turn our backs on women and children who left their home with nothing, nothing at all. That is not what America is supposed to be about. So I believe that working with Europe — and, by the way, you know, we've got some very wealthy countries there in that part of the world. You got Kuwait and you got Qatar and you got Saudi Arabia. They have a responsibility, as well. But I think this is a worldwide — that the entire world needs to come together to deal with

this horrific refugee crisis we're seeing from Syria and Afghanistan, as well. [applause]

Top 5 results for SANDERS

TOPIC: Gun Control

EXCERPT #1:

What we need to do is to do everything that we can to make certain that guns do not fall into the hands of people who do not have them. Now, I voted against this gun liability law because I was concerned that in rural areas all over this country, if a gun shop owner sells a weapon legally to somebody, and that person then goes out and kills somebody, I don't believe it is appropriate that that gun shop owner who just sold a legal weapon to be held accountable and be sued. But, what I do believe is when gun shop owners and others knowingly are selling weapons to people who should not have them — somebody walks in. They want thousands of rounds of ammunition, or they want a whole lot of guns, yes, that gun shop owner or that gun manufacturer should be held liable.

EXCERPT #2:

Let me just reiterate — just reiterate so there is no confusion, decades ago, before it was popular, in a rural state with no gun control, Bernie Sanders said, let's ban assault weapons, not see them distributed in the United States of America.

EXCERPT #3:

What I have said, is that gun manufacturer's liability bill has some good provisions among other things, we've prohibited ammunition that would've killed cops who had protection on. We have child safety protection work on guns in that legislation. And what we also said, "is a small mom and pop gun shop who sells a gun legally to somebody should not be held liable if somebody does something terrible with

that gun." So what I said is, " I would re-look at it." We are going to re-look at it and I will support stronger provisions.

EXCERPT #4:

A statement that was refuted by the governor of the state of Vermont, who was a supporter of hers, who said, yeah, in campaigns people tend to exaggerate. Here is the fact on guns. Let's talk about guns. That horrible, horrible Sandy Hook — what's the word we want to use, murder, assault, slaughter, unspeakable act. Back in 1988, I ran for the United States Congress one seat in the state of Vermont. I probably lost that election, which I lost by three points, because I was the only candidate running who said, you know what? We should ban assault weapons, not seen them sold or distributed in the United States of America. I've got a D-minus voting record from the NRA. [applause] And, in fact, because I come from a state which has virtually no gun control, I believe that I am the best qualified candidate to bring back together that consensus that is desperately needed in this country. [applause]

EXCERPT #5:

Well, I think Secretary Clinton knows that what she says is very disingenuous. I have a D-minus voting record from the NRA. I was in 1988, there were three candidates running for congress in the state of Vermont, I stood up to the gun lobby and came out and maintained the position that in this country we should not be selling military style assault weapons. I have supported from day one and instant background check to make certain that people who should have guns do not have guns. And that includes people of criminal backgrounds, people who are mentally unstable. I support what President Obama is doing in terms of trying to close the gun show loop holes and I think it should be a federal crime if people act as dormant. We have seen in this city a horrendous tragedy of a crazed person praying with people in the coming up and shooting nine people. This should not be a political issue. What we should be doing is working together. And by the way, as a senator from a rural state that has virtually no gun control, I believe that I am in an excellent position to bring people together to fight the sensible...

Top 5 results for CLINTON

TOPIC: Infrastructure

EXCERPT #1:

Well, I support what President Obama is doing. He called for and got accountability from the officials at the EPA to make sure the state is doing its job. He expanded Medicaid for helping kids particularly to get the health care they need. He is also ordered that there be a Head Start program. I support that. When it comes to the water itself, we are supporting a program that Mayor Waver announced through Flint Waterworks to pay people in Flint, not outsiders, but people here to deliver the water while we are fixing the pipes. I would do even more of that. As president, what we were able to put together was a beginning. As president, I would concentrate resources on the city for economic development for more jobs as we fix the water and provide the health and education, interventions that children need.

EXCERPT #2:

You know what - I have told my friends at the top of both unions, we've got take a look at this because it is one of the most common criticisms. We need to eliminate the criticism. You know, teachers do so much good, they are often working under most difficult circumstances. So anything that could be changed, I want them to look at it. I will be a good partner to make sure that whatever I can do as president, I will do to support the teachers of our country.

EXCERPT #3:

Well, there's no doubt, we have an enormous backlog of infrastructure repairs. Here's what I'd do - the congress finally got around to pass the Highway Transportation Bill, which should never have been partisan - it turned into be one partisan argument. I want to go further. I want to put 250 billion dollars additional, on top of what Congress has done. That gets us to a half a billion. I want to start a National Infrastructure Bank. I want to capitalize it with 25 billion that I believe will leverage 10 times that, that's another 250 billion. So

I'm trying to do this in a way that will gain support and be affordable but there's no doubt, we have to do more on our roads, our bridges, our tunnels, our ports, our airports. And as we've talked in the beginning here, under our ground, our water systems, our sewer systems. We have pipelines that are leaking and that are dangerous. We have so much work to be done and if we can put millions of peoples to work, I think my plan is a very good way to begin doing that work and get people out there doing it.

EXCERPT #4:

Well Nikki, it's a very fair question. Your government at all levels have let you and your children and the people of Flint down. So I think that there are there several things. All the repair work that is being done - and Mayor Weaver program that we support to begin to help train people in Flint to be able to do this work to distribute the water. Everything that is done has to be triple checked to regain your trust and to hold those who are responsible for fixing the pipes and delivering the clean water which has you said, you bathe in it and drink it and wash food in it. You do everything with it. Every one of us should have to run through in our minds how we use water every single day to understand the pressures and the real pain that families are going through. So I will make sure as President that I double and triple check. I will work with elected officials who I trust like your mayor and like your senators and member of Congress so that we can assure you that when it's fixed, you can trust it. You deserve nothing less.

EXCERPT #5:

Well, I agree completely. I want to go further though. I want us to have an absolute commitment to getting rid of lead wherever it is because it's not only in water systems, it's also in soil, and it's in lead paint that is found mostly in older homes. That's why 500,000 children today have lead — lead in their bodies. So, I want to do exactly what you said. We will commit to a priority to change the water systems, and we will commit within five years to remove lead from everywhere. We were making progress on this in the 1990's. I worked with then Senator Obama to get more money, more support to do more to remove lead.

Top 5 results for CLINTON

TOPIC: College Affordability

EXCERPT #1:

Well, let me address college affordability, because I have a plan that I think will really zero in on what the problems are. First, all the 40 million Americans who currently have student debt will be able to refinance their debt to a low interest rate. That will save thousands of dollars for people who are now struggling under this cumbersome, burdensome college debt. As a young student in Nevada said to me, the hardest thing about going to college should not be paying for it. So then we have to make it more affordable. How do we make it more affordable? My plan would enable anyone to go to a public college or university tuition free. You would not have to borrow money for tuition. But I do believe — and maybe it's because I worked when I went through college; I worked when I went through law school — I think it's important for everybody to have some part of getting this accomplished. That's why I call it a compact.

EXCERPT #2:

Well, congratulations on your education. And also on your plans for the future. And here's what I will do. We're going to refinance everyone's existing student debt, 40 million Americans have student debt. [applause] And right now, I go around asking people at my events if they know what their interest rate is and the interest rates literally go from like 8 percent to 14 percent. It's outrageous that at a time when interest rates have been historically low, people borrowing money to invest in their education are paying some of the highest interest rates around. And you can refinance your house to get a lower interest rate. You can refinance your car. Corporations can refinance their debt. Under my plan, you will be able to also lower your debt, move into a program to pay it back as a percentage of your income and more than that, my plan for debt-free tuition at public colleges and universities will eventually eliminate any student debt. But for people who have it, I'm going to put a date certain that after a certain number of years, you no longer have to pay anything. The government has to quit making money off of lending money to young people to get their education. [applause]

EXCERPT #3:

Kevin, if I could just jump in. I -- I believe that we should make community college free. We should have debt-free college if you go to a public college or university. You should not have to borrow a dime to pay tuition. I want to use pell grants to help defray the living expenses that often make a difference, whether a young person can stay in school or not. I disagree with free college for everybody. I don't think taxpayers should be paying to send Donald Trump's kids to college. I think it ought to be a compact -- families contribute, kids contribute. And together we make it possible for a new generation of young people to refinance their debt and not come out with debt in the future.

EXCERPT #4:

Well, I have what I call the new college compact. Because I think everybody has to have some skin in this game, you know. Number one, States have been dis-investing in higher education. In fact, I think New Hampshire, in state tuition for public colleges and universities, is among the highest if not the highest in the country. So states over a period of decades have put their money elsewhere; into prisons, into highways, into things other than higher education. So under my compact, the federal government will match money that the states begin to put back in to the higher education system. Secondly, I don't believe in free tuition for everybody. I believe we should focus on middle-class families, working families, and poor kids who have the ambition and the talent to go to college and get ahead. So I have proposed debt free tuition, which I think is affordable and I would move a lot of the Pell Grant and other aid into the arena where it could be used for living expense. So I put all of this together, again, on my website and I've gotten such a good response. But I want to quickly say, one of the areas that Senator Sanders touched on in talking about education and certainly talking about health care is his commitment to really changing the systems. Free college, a single payer system for health--and it's been estimated we're looking at 18 to \$20 trillion, about a 40 percent in the federal budget. And I have looked at his proposed plans for health care for example, and it really does transfer every bit of our health care system including private health care, to the states to have the states run. And I think we've got to be really thoughtful about how we're going to afford what we proposed, which is why everything that I have proposed I will tell you exactly how I'm going to pay for it; including college.

EXCERPT #5:

Well, I have actually documented every way that I'm going to pay for what I'm doing because I think the American public deserves to know. And, you can go to my website and actually see that. But, there are serious questions about how we're going to pay for what we want to see our country do. And, I'm the only candidate standing here tonight who has said I will not raise taxes on the middle class. I want to raise incomes, not taxes, and I'm going to do everything I can to make sure that the wealthy pay for debt free tuition, for child care, for paid family leave. To help us bring down student debt we're going to refinance that student debt, saving kids thousands of dollars. Yeah, and that will also come out of the — some of the pockets of people in the financial services industry...

Top 5 results for SANDERS

TOPIC: Immigration

EXCERPT #1:

The answer is yes. We've got 11 million undocumented people in this country. I have talked to some of the young kids with tears rolling down their cheeks, are scared to death that today they may or their parents may be deported. I believe that we have got to pass comprehensive immigration reform, something that I strongly supported. I believe that we have got to move toward a path toward citizenship. I agree with President Obama who used executive orders to protect families because the Congress, the House was unable or refused to act. And in fact I would go further. What would motivate me and what would be the guiding light for me in terms of immigration reform, Judy, is to bring families together, not divide them up. And let me say this also. Somebody who is very fond of the president, agrees with him most of the time, I disagree with his recent deportation policies. And I would not support those. Bottom line is a path towards citizenship for 11 million undocumented people, if Congress doesn't do the right thing, we use the executive orders of the president. [applause]

EXCERPT #2:

I know that. But point being is that progressives did vote against that for that reason. My view right now — and always has been — is that when you have 11 million undocumented people in this country, we need comprehensive immigration reform, we need a path toward citizenship, we need to take people out of the shadows.

EXCERPT #3:

Poverty is increasing. And if wages are going down, I don't know why we need millions of people to be coming into this country as guestworkers who will work for lower wages than American workers and drive wages down even lower than they are right now. [end video clip] [crosstalk]

EXCERPT #4:

Well, when we talk about efforts to assist immigrants, Secretary Clinton prevailed upon the governor of New York, Elliot Spitzer, who wanted to do the right thing and provide driver's license to these who were undocumented, she said don't do it, and New York State still does not do it. In Vermont, by the way, I worked with officials and undocumented people in Vermont do have the ability to get driver's license. [applause] When we talk about immigration, the secretary will remember that one of the great tragedies, human tragedies of recent years is children came from Honduras where there's probably more violence than almost any place in this country, and they came into this country. And I said welcome those children into this country, Secretary Clinton said send them back. That's a difference.

EXCERPT #5:

A couple of years ago, I had the opportunity to go on a congressional delegation. And I went to one of these Turkish refugee camps right on the border of Syria. And what a sad sight that was: Men, women, children forced out of their homes. And Turkey, by the way, did a very decent thing, providing what was reasonable housing and conditions for those people. It seems to me that given our history as a nation that

Information Retrieval, Spring 2016 Tom Catullo & Vamsi Chunduru

has been a beacon of hope for the oppressed, for the downtrodden, that I very strongly disagree with those Republican candidates who say, you know what, we've got to turn our backs on women and children who left their home with nothing, nothing at all. That is not what America is supposed to be about. So I believe that working with Europe — and, by the way, you know, we've got some very wealthy countries there in that part of the world. You got Kuwait and you got Qatar and you got Saudi Arabia. They have a responsibility, as well. But I think this is a worldwide — that the entire world needs to come together to deal with this horrific refugee crisis we're seeing from Syria and Afghanistan, as well. [applause]