

Neural Argument Generation Augmented with Externally Retrieved Evidence (Supplementary Material)

Xinyu Hua and Lu Wang

College of Computer and Information Science

Northeastern University

Boston, MA 02115

hua.x@husky.neu.edu luwang@ccs.neu.edu

1 Data Preprocessing

Filtering Reddit Threads. The following list of keywords are used to match the discussion thread title, where we remove a thread from our dataset if its title contains any of the keywords:

upvote, downvote, reddit, subreddit, karma, delta.

Bootstrapping Algorithm for Building Domain Classifier. As described in Section 3 of the paper, we aim to use a domain classifier to identify discussion threads related to politics. We leverage the Wikipedia abstracts to automatically construct the training set. We first collect 1,285,705 abstracts from DBPedia.¹ Then we manually create two lists of politics keywords and non-politics keywords, as shown in Table 1. Based on the keywords, we automatically label Wikipedia abstracts as politics or non-politics by matching keywords with their titles. The resultant dataset contains 264,670 positive samples and 827,437 negative samples. A logistic regression model is trained on this dataset with unigram TF-IDF scores as features.

Initially, we observe a good amount of false negative results when tested on CMV posts. We thus continue training the model by adding training samples from CMV that are classified with high confidence. This procedure stops after three iterations when reasonably good performance is observed based on human inspection. The model labels 12,549 threads as politics and 13,731 as non-politics.

2 Experiments

2.1 Effect of Pre-training

As described in Section 6.2 of the paper, during training time we initialize LSTM parameters with a pre-trained seq2seq model based on OP

¹<http://dbpedia.org/page/>

Politics		non-Politics	
politics	political	science	media
policy	congress	automobiles	sports
rights	election	football	fashion
president	trump	entertainment	movie
clinton	immigration	movies	music
democracy	democrats	musics	art
democratic	republican	arts	television
constitution	liberal	religion	philosophy
government	legalization	morality	dating
surveillance	amnesty	eugenics	marriage
antisemitism	terrorism	parenthood	history
war	taxation	organic	handicaps
liberalism	libertarianism	disease	
marxism	conservatism		
anarchism	autocracy		
fascism	voting		

Table 1: Lists of politics and non-politics keywords used to obtain Wikipedia abstract labels.

and argument pairs without evidence. Here we show the effect of pre-training by comparing ME-TEOR scores of models with and without pre-training. We test on our separate decoder model and shared decoder model, both with attention over keyphrases. A seq2seq baseline is shown for comparison. From Figure 1 we can see that pre-training boosts performance for all models.

2.2 Topic Relevance Evaluation Model

We evaluate topic relevance of generated argument based on a relevance estimation model. Our relevance estimation model is inspired by the latent semantic model described in Huang et al. (2013), which projects queries and web-documents onto a common low-dimensional space. Then the relevance can be estimated by the dot-product of projected vectors. Our implementation is simplified, where OP or argument is represented as the average of the 300-dimensional GloVe (Pennington et al., 2014) word embeddings. Vectors of OP and argument are then fed into two layers of linear transformations and projected onto a 200-dimensional space. We take the dot-product of

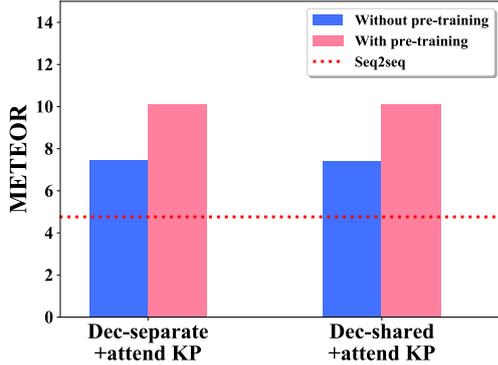


Figure 1: Effect of pre-training on seq2seq-based models. Red dotted line shows the performance of the seq2seq model used for parameter initialization.

Model	MRR	P@1
Our Relevance Estimation Model	95.30	92.09
Random	37.76	16.82

Table 2: Ranking results for our trained relevance estimation model and random ordering.

these transformed representations for OP and argument, and feed into a sigmoid function to produce the relevance probability.

Model Training and Test. We reuse our training data, and retain samples with argument length between 50 and 140 tokens. We further divide it into training (28,408), development (3,514), and test (3,703) sets. For each OP and argument pair, we first randomly sample 100 arguments from other OPs. Then we pick the top 5 dissimilar ones to the given argument, measured by Jaccard distance, as negative training examples. Cross-entropy loss is used as training objective:

$$\mathcal{J}(\theta) = -\frac{1}{|D|} \sum_{(\mathbf{x}_1, \mathbf{x}_2, z) \in D} \log P(z|\mathbf{x}_1, \mathbf{x}_2; \theta) \quad (1)$$

$\mathbf{x}_1, \mathbf{x}_2$ are averaged word embedding representations for OP and argument, z is the binary label, and θ is the model parameter. We optimize the training objective using Adam (Kingma and Ba, 2015) with 0.001 as the initial learning rate.

During test, the model will predict relevance score for both positive samples and negative samples. We then rank these samples according to the score, and compute average Mean Reciprocal Rank (MRR) and Precision at 1 (P@1) for evaluation. Results on the test set is reported in Table 2.

For comparison, we also show a random ordering baseline.

3 Sample System Arguments

Additional sample arguments output by different systems are displayed in Table 3 and Table 4 along with generated keyphrases by our models.

Input (Original Post):

Title: CMV: The idea that the USA cannot “afford” universal socialised medicine makes no sense

Elaboration: Whenever the discussion of socialised healthcare in the US comes up, I always see one central argument against it. It’s the idea that Americans, and America as a whole, simply can’t afford to pay for such a wide and expansive government programme. That it would place a massive strain on the economy and that overall it would cost too much.

I don’t understand where this idea comes from, because from an economic point of view it makes no sense. The US spends approximately 17.2% of its GDP on Healthcare, as compared to less than 10% seen in countries with smaller economies that have socialised universal healthcare coverage.

I also can’t see how it would cost Americans individually more. If all Americans are paying into a single healthcare provider (in the form of tax) then surely the costs for each individual would go down as compared to dealing with dozens of individual health insurance providers. This way the costs would be more easily shared. At the very least, it couldn’t possibly cost the individual American more money than insurance or other healthcare provision costs now.

I’m looking to understand this idea. Is there a genuine reason why people think the US can’t afford socialised healthcare? I understand there could be ideological reasons why people object, but I’m looking to understand the economics. Is there a good reason that the US couldn’t afford universal healthcare.

For perspective, I’m from the UK, and we have socialised medicine here, in the form of the National Health Service (NHS). This means that aside from a very small number of privately operated hospitals, we can go to any hospital anywhere, and get treatment. This is free at the point of demand, and paid for by the government via taxes and government borrowing.

Human: The laws of supply and demand still rule our nation. The problem is that whenever government adds taxpayer money into the equation it skews the demand side. more demand than supply causes market forces to push costs per unit up.

Retrieval: According to 2009 World Bank statistics, the U.S. had the highest healthcare costs relative to the size of the economy (GDP) in the world, even though estimated 50.2 million citizens (approximately 15.6% of the September 2011 estimated population of 312 million) lacked insurance. Clinical trial management: bio-pharma companies (sponsors) are required to conduct extensive clinical trials to get their new drugs approved for launch by FDA and other global regulatory agencies such as the European Medicines Agency (EMA) and the Pharmaceuticals and Medical Devices Agency (PMDA) of Japan. The cost of insurance has been a primary motivation in the reform of the US healthcare system, and many different explanations have been proposed in the reasons for high insurance costs and how to remedy them. In this sense the Canadian healthcare system is merely a publicly funded one where services are provided by a mixture of public and private entities, which most Canadians appreciate and desire. In the 1960s, there was a plethora of public initiatives by the federal government to consolidate and modernize the U.S. healthcare system.

Our Model with Dec-separate+attend KP: I’m sorry, this is following [this cmv] ([URL]) . for medical reasoning, we need to care about the quality of Americans. Everybody else has a high risk on public track. There is no reason to think they know healthcare. However, it isn’t impossible to take a lot of high standards, they’re economic.

Generated KP: the highest costs; world bank; the world; medical expense; the OECD hospital system; medical and other nations; the Canadian healthcare system; public and private authorities; the [UNK] VA medical center.

Our Model with Dec-shared+attend KP: Legalized greed means the implementation of health insurance is rising. I don’t really care what you’re getting. I don’t really care what you do. I don’t really disagree with your heart.

Generated KP: a heart and heart disease; health insurance.

Seq2seq + encode evd: 1. This is a lot of money. If you don’t care about the cost of spending a lot of dollars, it would be more difficult for the rest of the tax of the economy. The rest of the welfare of dollars in terms of dollars. The cost of dollars to pay taxes. The cost of the government isn’t going to be able to make the cost of the economy.

Table 3: Comparison of human written arguments and our system generated arguments.

Input (Original Post):

Title: CMV: The United States should drastically increase foreign aid.

Elaboration: I am operating under the assumption that the only goal of a moral country should be to decrease suffering and increase happiness.

22,000 children die each day due to poverty related causes (such as malnutrition, starvation, lack of clean water, malaria, TB, etc). These lives are neither difficult nor expensive to save. The Against Malaria Foundation saves one life for every 3,000 dollars donated. Our national budget is 3.9 trillion; we spend 29 billion on foreign aid (less than 1%)- which is not even aimed at saving lives, reducing suffering and increasing happiness in the most efficient way.

I am not arguing that countries shouldn't focus on their own citizens; of course we need to make sure Americans are taken care of. But we could cut things such as eliminating excess military spending, which does not drastically increase happiness or reduce suffering, and we could change the lives of millions of people.

My view basically comes down to this; suffering, pain, and loss, are just as horrible for people in the third world to experience as Americans. We can either try to slightly improve the lives of Americans who are already well off, or we can end the unimaginable suffering of those in extreme poverty. Help me understand why a world where the average quality of life is higher is worse than one where it is lower.

If you argue that the money will just be taken by corrupt governments, I would argue that we can focus on NGO's where this is not a problem. With our resources, knowledge, and connections, we certainly could figure out a way to do good things with are money. The other argument that I can think of is that countries need to find a way to lift themselves out of poverty. Political scientists and historians do not see sub saharan Africa as having the resources, institutions, or infrastructure to do this anytime soon. Instead, we can help them establish their economies by helping with things such as education.

I believe that foreign aid is so low because it is not politically convenient and we are not focused on doing what is morally right. Is there another legitimate reason?

Human: Let's say that instead of giving money to many countries we attempted to create a place where investors and the country can take what resources the country has and parlay that into some sort of industrial revolution. Give a man a fish, teach a man to fish.

Retrieval: They agree that bad policy is detrimental to economic growth, which is a key component of poverty reduction, but have found that aid dollars do not significantly incentivize governments to change policy. Also in 1948, the United States and the recipient countries created the Organisation for European Economic Cooperation (OEEC it became the OECD in 1961) to coordinate the use of the aid. In 2002, total gross foreign aid to all developing countries was \$76 billion. Al-Hussain, A. and Johnson, R. (2009) found a strong relationship between the efficiency of corporate governance structure and Saudi bank performance when using return on assets as a performance measure with one exception - that government and local ownership groups were not significant. In cases where people lack representation by a state's government, the unrepresented become a separate people. While government aid is generally called foreign aid, aid that originates in institutions of a religious nature is often termed faith-based foreign aid. Because faith-based foreign aid focuses on churches or organizations operating independently of the government, funding has a better chance of being used effectively.

Our Model with Dec-separate+attend KP: We are moving toward our economy to do so much worse. We are moving toward our economy. However, we need a lot more dollars than other things . [URL] there are benefits [URL] [URL] [URL] global warming works. [URL] global warming will place increasing money from individuals. [URL] global warming [URL] [URL] [URL] there's low individuals for individuals, and can't convince 80 people own them. [URL] people on the US.

Generated KP: the United States; foreign index; Tax; other countries; the United Kingdom; foreign countries; the United States government; higher education system in the United States government; national security and nations; their economic care; the national transportation authority; large societies.

Our Model with Dec-shared+attend KP: Increasing charities for people who don't get money in the middle east etc). It's not a bad thing. The problem is that people don't have a problem with the world. The US doesn't have a problem with people who don't live in the US.

Generated KP: developing countries of developing countries; the country; a country.

Seq2seq + encode evd: 1. The only reason why the problem is that they need to be able to improve their lives. They don't know how much they have to deal with the rest of the lives of the lives of their lives. The rest of the world, they have a lot of money on the rest 99 %. The rest of the world, I'd like to start with the rest of the world they get back on their dollar.

Table 4: Comparison of human written arguments and our system generated arguments.

References

- Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, pages 2333–2338.
- Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.