# POLITICAL IDEOLOGY DETECTION USING RECURSIVE NEURAL NETWORKS

ACL (Association for Computational Linguistics) 2014,

University of Maryland

Presented By: Navita Jain

# IDEA:

- An individual's words often reveal their political ideology.

- How do we know a sentence contains Ideological bias?

  If author's political position (liberal or conservative : "US politics)  is evident

  from the text.

# TASK:

- Given a sentence, detect the political ideology expressed in the sentence.

- In this paper authors model a political ideology under two category:
  - liberals (democrats) and conservatives (republican)
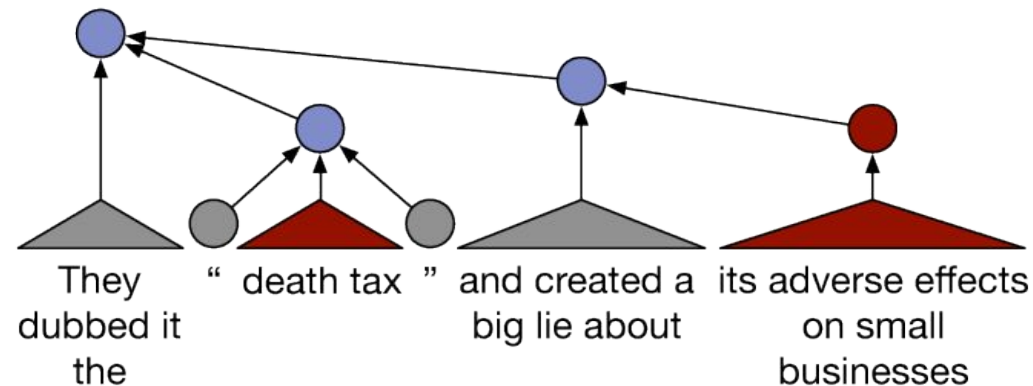  - Even though not true, authors models it this way.

# PREVIOUS WORK:

- A traditional approach to this problem is bag of words model
    - Each word treated as separate feature.

Issue: syntactic structure is ignored. We will see how syntactic structure is important.

# DETECTING POLITICAL IDEOLOGY IS CHALLENGING:

▪ Not just from system point of view but also from humans stand point:

  ▪ Requires Political knowledge
    _ phrase "death tax" and "small businesses" -> conservatives

  ▪ Ability to identify subtle elements of language use
    _negative propositional attitude "big lie about" is indicative of that author is not conservative but -> liberal



```
They          " death tax "   and created a   its adverse effects
dubbed it                     big lie about   on small
the                                           businesses
```

while certain sections of the sentence are locally conservative, the way they are used in context makes the overall sentence liberal.

# RECURSIVE NEURAL NETWORK

- A  supervised machine learning tool, that uses a parse tree/ hierarchal structure of language to capture the syntactic and semantic composition of a sentence.

- RNN models semantic composition:
  - The idea is that meaning of a sentence or text can be decomposed into subparts.
  - RNN models this by recursively combining vector representation of words and phrases.

  - Word vectors are used as features and serve as the basis of sequential classification. They are then grouped into subphrases, and the subphrases are combined into a sentence that can be classified by sentiment and other metrics.

# RNN CONTINUED..

phrase        d dimensional vector

d dimension        d dimension

words

- Each word is modeled as a vector of size d.

- A network is constructed that take
  - Input :  2 vectors concatenated into vector of length 2d.
  - Outputs : A vector of length d.

- RNN stores this information and then uses it as input to some other network.

- This way we can combine multiple words to phrases and phrases to sentences. We don't need to have a separate network for each of these levels, and can just reuse the same network each time _ which is why these are called *recursive neural networks*.
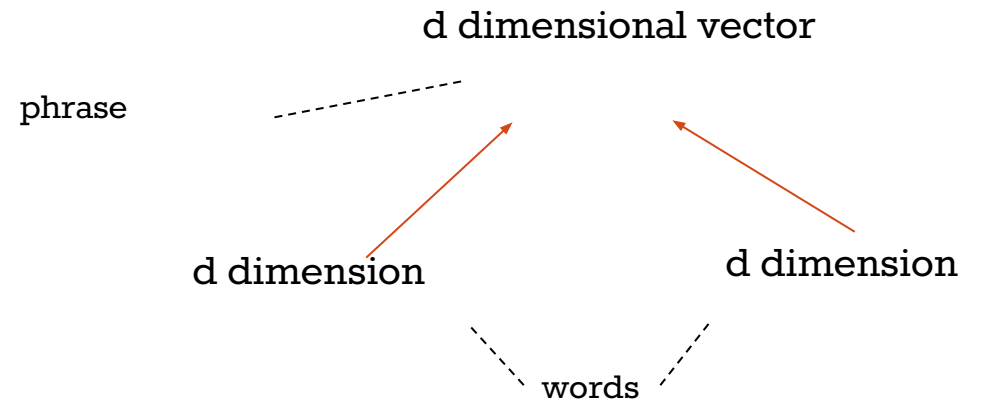
- Basic Idea: supervised learning

  Each word in sentence/phrase is a vector.

  Say, phrase is **"so-called climate change"**

  $x_a$ : climate

  $x_b$ : change

  word vector

  Words form phrases (based on phrase tree), each phrase is also a vector
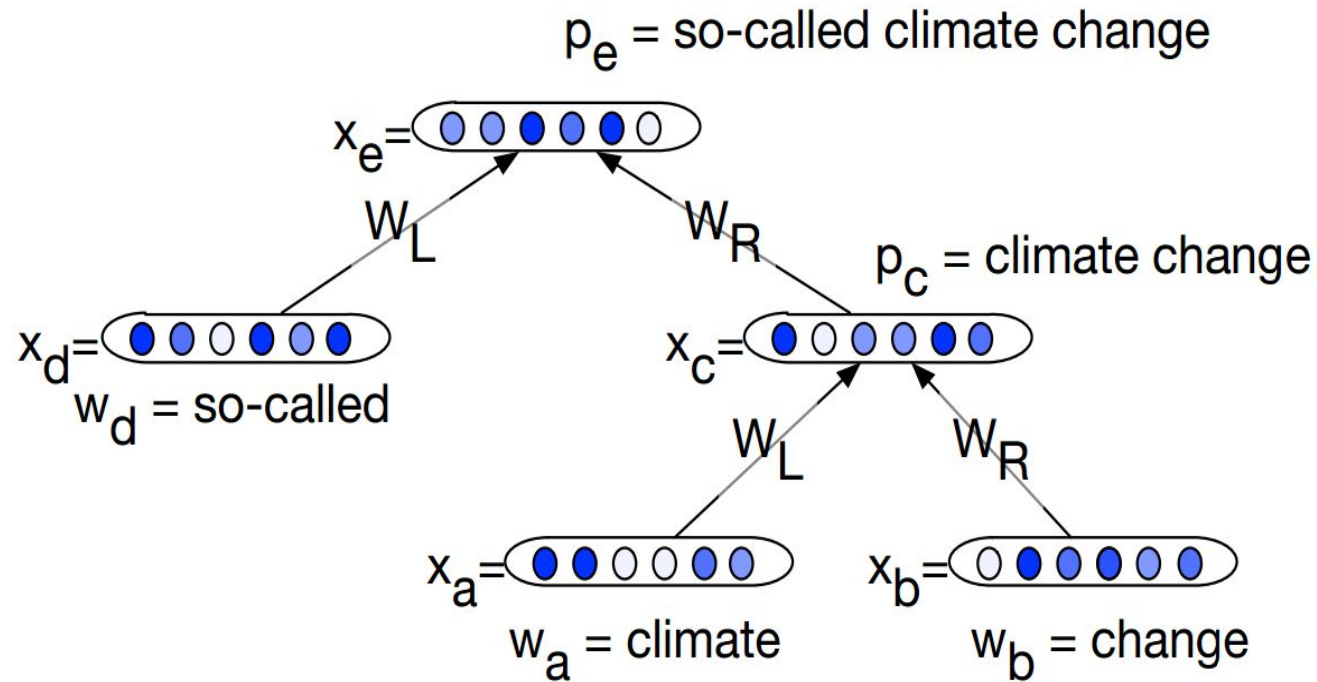
  $x_c$ : climate change

  $x_d$ : so-called

  phrase vector

  These phrases combine to form complete sentence(representation), such that   meaning of whole sentence is retained.

  complete representation or another phrase vector

  $x_e$ : so-called climate change

$p_e$ = so-called climate change

$x_e=$ ⬭(○○●○●○)

$W_L$　　　$W_R$

$p_c$ = climate change

$x_d=$ ⬭(●●○●○●) $x_c=$ ⬭(●○●●●●)

$w_d$ = so-called

$W_L$　　　$W_R$

$x_a=$ ⬭(●●○○●●) $x_b=$ ⬭(○●●●●●)

$w_a$ = climate　　　$w_b$ = change

for example: words $w_a$ and $w_b$ are merge to form phrase p

i.e.　vector $x_a$ and $x_b$ are concatenated to form phrase vector $x_c$

$$x_c \;=\; f(\, W_L \cdot x_a + W_R \cdot x_b + b\,)$$

$W_L$ and $W_R$ are some composition matrix for left positioning of word and right positioning of word.

The first step towards an RNN is vectorization
　　Two ways to model vectors and matrices
　1.　Randomily initialize the matrices

　2.　Word2Vec: Is an algorithm that converts a corpus of words into vectors, which can then be thrown into a vector space to measure the cosine distance between them; i.e. their similarity or lack of similarity.

- Since this is more of a classification task. On the output of this function they used a model called Softmax, which is simply a way to convert raw scores to probabilities. This help them in labeling a sentence into liberal or conservative.

- They then trained their model on Cross-entropy loss , to minimize the discrepancy in their categorization and ground truth.

- The idea here is that if an element is associated with liberal bias, its vector space should be distinct from vector of element indicative of  conservative bias

# DATASET AND EVALUATION

- Evaluated on two datasets -
  - Convote Dataset: Consist of US Congressional floor debate transcripts. These dataset consist of 7816 sentences, half of which were from Democrats and other half from Republicans.

  - Ideological Books Corpus (IBC): This consist of articles from books and magazine written by authors whose political bias is known.
    - Author extended the dataset by providing sentence-level and phrase level annotation, crowd-sourced through Crowdflower. After filtering, the dataset consisted of 3412 sentences.

  - The authors apply filters on both the datasets to only keep sentences containing explicit bias in either direction, and remove neutral sentences

Proportion of liberal, conservative and neutral annotations with respect to node depth (distance from root)



Label Probability vs. Node Depth
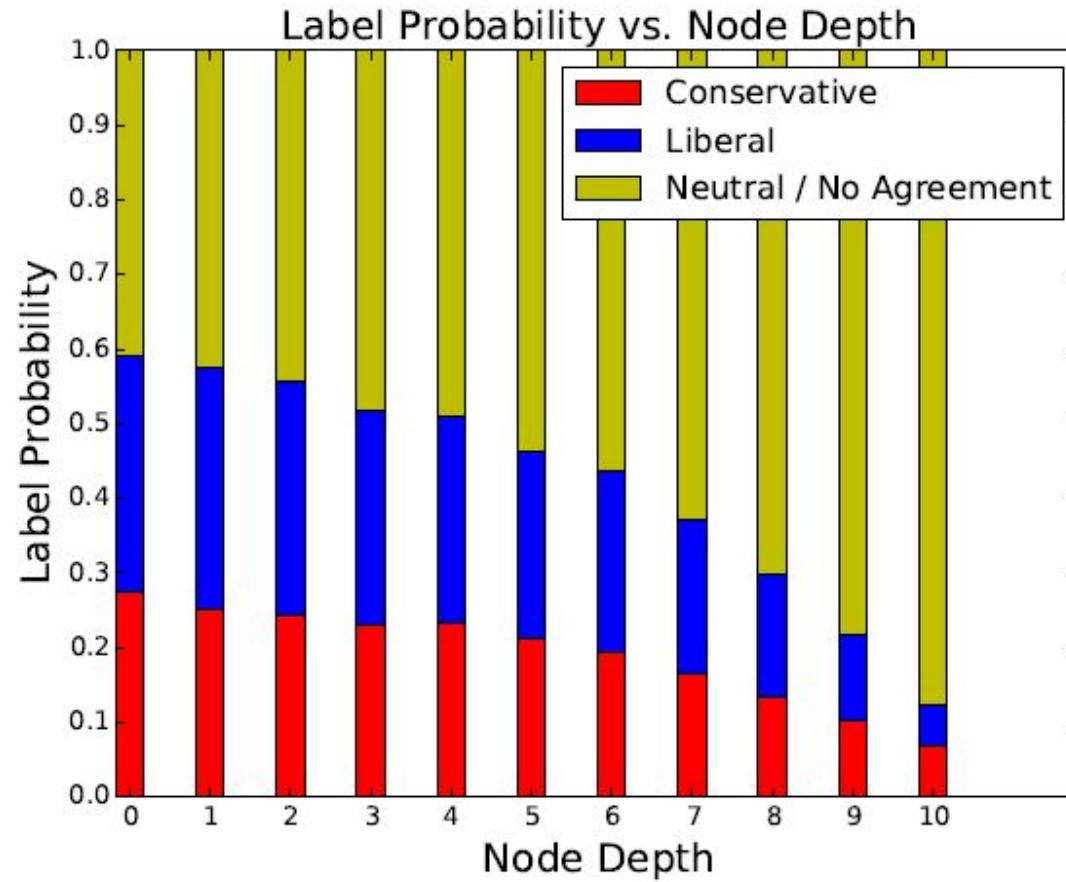
Lower-level phrases are almost always neutral while full sentences are much more likely to be biased. So they stayed with task of identifying sentence as liberal or conservative.

# PERFORMANCE EVALUATION

Baseline:

- Random : the label (liberal/conservative), chosen        randomly
- LR1 : logistic regression using bag-of-words feature
- LR2 : logistic regression using bag-of-words and       phrase annotation in training data
- LR3 : logistic regression _ bag-of-words and dependency based features
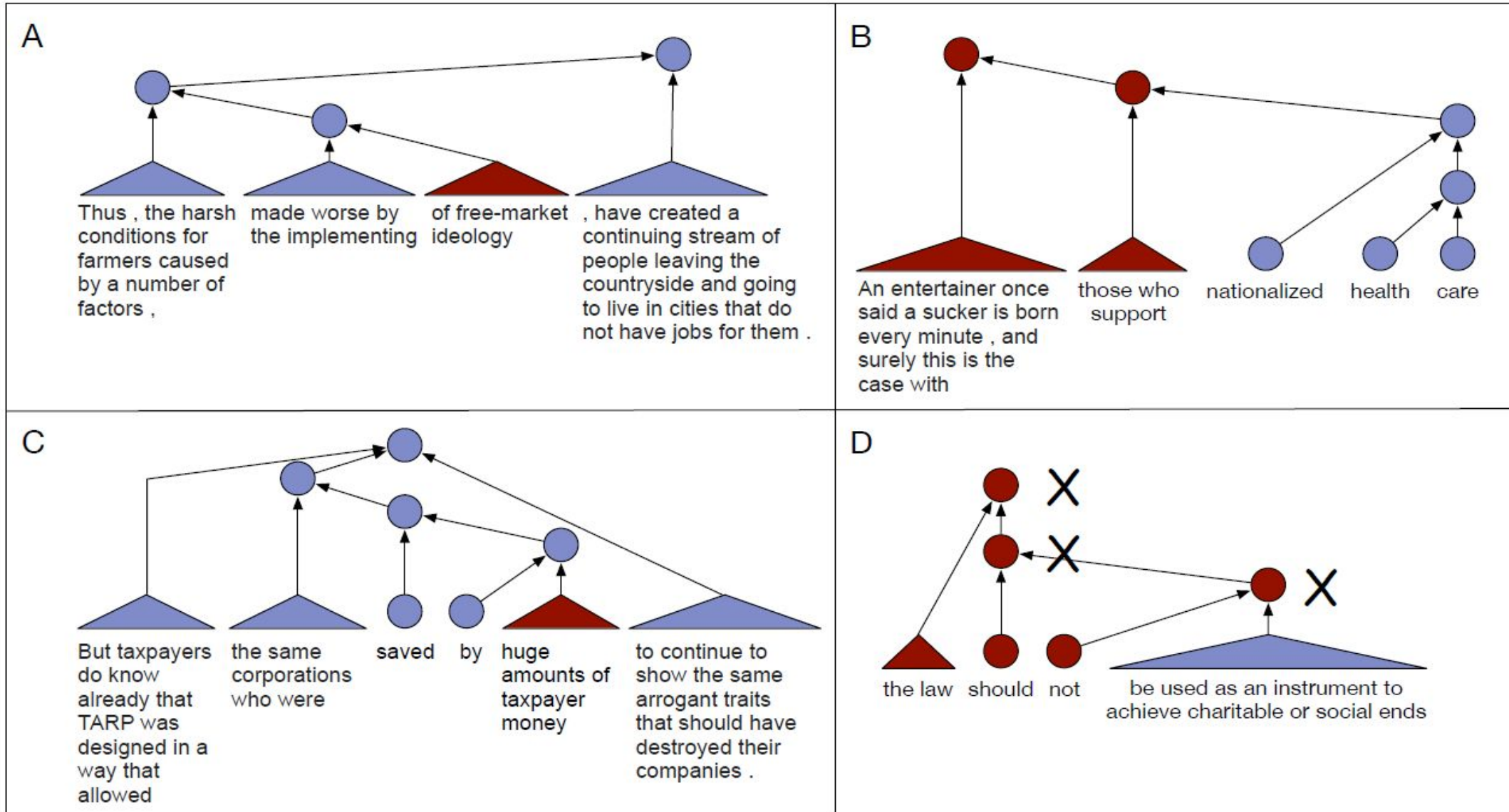- LR-(w2v) : logistic regression over avg word vectors

RNN configurations:

-  RNN1 : basic RNN trained on sentence annotation and random initialization
-   RNN1-(w2v) : as above only initialization with word2vec
-   RNN2-(w2v) : trained on sentence and phrase-level annotations and initialized using word2vector.

Results on Convote better than on IBC:

    Explanation :
-   Large dataset of Convote (more data better training)
-   Shorter sentences in Convote.

# QUALITATIVE ANALYSIS



**A**

Thus , the harsh conditions for farmers caused by a number of factors ,

made worse by the implementing

of free-market ideology

, have created a continuing stream of people leaving the countryside and going to live in cities that do not have jobs for them .

**B**

An entertainer once said a sucker is born every minute , and surely this is the case with

those who support

nationalized    health    care

**C**

But taxpayers do know already that TARP was designed in a way that allowed

the same corporations who were

saved    by    huge amounts of taxpayer money

to continue to show the same arrogant traits that should have destroyed their companies .

**D**

the law   should   not

be used as an instrument to achieve charitable or social ends

Highest probability n-grams for conservative and liberal ideology

# DISCUSSION

- 1. What the performance difference would it make if the root level predictions from the RNN were used directly for evaluation?
- 2. How could the Authors have included a wider spectrum of political ideologies as not everyone is liberal on every issue and vice versa.
- 3. How does label imbalance in the sentences in a test set affect the accuracy of predicting ideological bias?
- 4. Author mentions that semantic composition does not apply to sarcasm and idioms, so what steps can be taken to predict ideological bias in sarcastic sentences and idioms?
- 5. Could a similar approach be used to gauge public opinion from a collection of posts?
- 6. while recent work in sentiment analysis that successfully models the compositional aspect of language is inspirational ,is there any approach model yet to account language for which semantic composition does not apply like sarcasm and idioms
- 7. How can the model be modified to detect political ideologies from social networking websites like twitter, facebook, etc?
- 8. How does negation confuses the RNN? Isn't there a way to teach the RNN to handle negation
- properly?

# THANKS

Other references:

http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

http://www.marekrei.com/blog/political-ideology-detection/

http://deeplearning4j.org/recursiveneuraltensornetwork.html

**the Republican leadership**

- ◉ Neutral
- ○ Conservative
- ○ Liberal
- ○ Not neutral, but I'm unsure of which direction

**the Republican leadership making clear it wanted no piece of meaningful health care reform**

- ○ Neutral
- ○ Conservative
- ◉ Liberal
- ○ Not neutral, but I'm unsure of which direction

**But , with the Republican leadership making clear it wanted no piece of meaningful health care reform , few Republicans were interested in nego-tiating seriously .**

- ○ Neutral
- ○ Conservative
- ◉ Liberal
- ○ Not neutral, but I'm unsure of which direction

Annotation examples used in crowdsourcing platform : Crowrdflower to gather political ideology annotations. showing incremental reveal of progressively longer phrases.

An individual's words often reveal their political ideology.

Problem:

Detecting ideological bias on the sentence level

Approach:

Identify political position evinced by sentence using Recursive Neural Network framework/technique

How do we know a sentence contains Ideological bias?

If author's political position(liberal or conservative : "US politics) is evident from the text.

# WHY IS IT CHALLENGING:

- Well detecting ideological bias is a challenging task to even humans because :
  - Requires political knowledge
    - ‒ phrase "death tax" and "small businesses" -> conservatives
  - Annotators ability to identify subtle elements of language use
    - ‒ However negative propositional attitude "big lie about" is indicative of that author is not conservative but -> liberal



They dubbed it the " death tax " and created a big lie about its adverse effects on small businesses
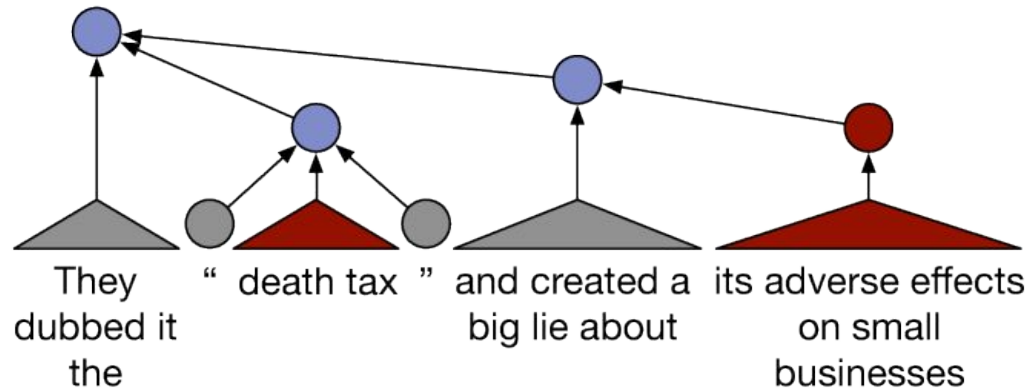
# RELATED WORK

- RNNs are *quantitatively* more effective than existing methods that use syntactic and semantic features separately, and we also illustrate how our model correctly identifies ideological bias in complex syntactic constructions.

# RECURSIVE NEURAL NETWORK:

▪ Basic Idea: supervised learning

Each word in sentence/phrase is a vector.

Say, phrase is **"so-called climate change"**

$x_a$ : climate

$x_b$ : change

word vector

Words form phrases (based on phrase tree), each phrase is also a vector

$x_c$ : climate change

$x_d$ : so-called

phrase vector

These phrases combine to form complete sentence(representation), such that   meaning of whole sentence is retained.

complete representation or another phrase vector

$x_e$ : so-called climate change

# RECURSIVE NEURAL NETWORK

- RNN is a machine learning technique where

# CHALLENEGE: HOW TO COMBINE VECTORS, SO AS TO RETAIN THE WHOLE MEANING OF SENTENCE??

# LEARN REPRESENTATION THAT CAN DISTINGUISH POLITICAL POLARITY GIVEN LABELED DATA

- If element represents liberal bias, its vector should be distinct from vector of conservative bias element.

- RNN achieve this distinction, by make a prediction
  - Prediction should match annotated data.
  - If not the case (prediction does not match annotation):
    - Discrepancy measured through cross-entropy loss.