Natural Language Processing – Homework #1

Xingyu He xh2513@columbia.edu

Oct 6 2021

---

1 Written Portion

1. Experimental Design (5 points) Describe how you set up your experiments to investigate what sets of features are useful for predicting the outcome of debates on religious topics vs. debates on all other topics.

   (a) I partition the debate data into religious debates and non-religious debates. Within the same type of features excluding ngrams (e.g. lexicon, linguistic, and user features), I iterate through possible features and select the best performing 2 lexicon and 2 linguistic features in terms of accuracy on the validation set via cross validation. These sets of models are run on both religious debates and all other debates. By comparing and contrast the set of features that yield the highest accuracy, we will hopefully gain insights as to which sets of features are useful for predict the outcome of debates on religious topics and non-religious topics.

2. Performance. (5 points) Show the accuracy for all models, including a simple majority baseline. Show all results in a single table.

   See end

3. Analysis (20 points)

   (a) What two linguistic features did you incorporate in the Ngram+Lex+Ling model? State your hypothesis about which linguistic features you think would be useful (and why), and perform some data analysis to show that they would be useful to incorporate into your model. Give two examples where these features would help you identify the winner of a debate.

      i. The two linguistic features I incorporate in the Ngram+lex+Ling model is the number of swear words in a given debate on each side and the number of times one side reference the opponent, either by the opponent's user name or the word opponent.

      ii. For the number of swear words, I hypothesize that debater who swears a lot may be perceived as having ill intention to the debate, which may push voters to side with the other side. For the number of references to opponents, I hypothesize that the debater who addresses their opponent more tend address their opponent's arguments. This is better than if the debater leaves their opponent's arguments unaddressed.

      iii. The table below shows the best performing accuracy across all other features for all possible linguistic features of size 2. In other words, each row represents the highest accuracy on the validation set for models that includes the corresponding linguistic features. Overall, the accuracies on the validation set do not differ by that much, presumably because the tri-gram model can already explains a lot of the linguistic features here. However, when the feature set includes the opponent reference counts and the swear counts, the accuracy on the validation set is higher than 4 percentage points than the model with the lowest accuracy in the table. However, since there only 1,592 debates in the training set and 399 debates in the validation set, the different is not significant enough. One that is worth noting is that swear words seem to play a more significant role than opponent reference counts. The top 8 best models in the table all have swear words as a linguistic features.

linguistic features	val accuracy	train accuracy
['reference_to_opponent', 'swear_words']	0.769424	0.750628
['swear_words', 'personal_pronouns']	0.764411	0.754397
['swear_words', 'number']	0.754386	0.742462
['swear_words', 'questions']	0.754386	0.736809
['reference_to_opponent', 'websites']	0.751880	0.750628
['length', 'swear_words']	0.751880	0.744975
['swear_words', 'modal_verb']	0.751880	0.744347
['personal_pronouns', 'number']	0.749373	0.744975
['websites', 'exclamation']	0.749373	0.744347
['length', 'reference_to_opponent']	0.749373	0.748744
['reference_to_opponent', 'personal_pronouns']	0.746867	0.745603
['reference_to_opponent', 'number']	0.746867	0.744975
['personal_pronouns', 'questions']	0.746867	0.750000
['personal_pronouns', 'exclamation']	0.746867	0.736809
['length', 'personal_pronouns']	0.746867	0.734925
['websites', 'number']	0.744361	0.744347
['swear_words', 'websites']	0.741855	0.748116
['questions', 'modal_verb']	0.741855	0.728015
['questions', 'websites']	0.741855	0.743090
['reference_to_opponent', 'modal_verb']	0.741855	0.736809
['length', 'number']	0.741855	0.744347
['reference_to_opponent', 'questions']	0.741855	0.742462
['websites', 'modal_verb']	0.741855	0.746859
['length', 'websites']	0.736842	0.748116
['reference_to_opponent', 'exclamation']	0.736842	0.733040
['personal_pronouns', 'modal_verb']	0.734336	0.750628
['personal_pronouns', 'websites']	0.734336	0.757538
['length', 'modal_verb']	0.731830	0.741206
['questions', 'number']	0.729323	0.734925
['swear_words', 'exclamation']	0.729323	0.724874
['number', 'modal_verb']	0.729323	0.726131
['length', 'questions']	0.729323	0.734296
['length', 'exclamation']	0.729323	0.744347
['exclamation', 'number']	0.726817	0.724246
['questions', 'exclamation']	0.726817	0.724874
['exclamation', 'modal_verb']	0.721805	0.721734

iv. An example is the following debate titled "from an official capacity, the catholic church has never contradicted itself". The con side in the debate initiated an attack on the pro side with foul language. For example, the con side said "Your stupid for thinking the catholic church has not had a contradiction in 2000 years. for starters the catholic priests are called fathers ex. the name father john. when it clearly states that no one should use the father for he is in heaven and that statement was made by a priest." On the other hand, the pro side remains civil without using any swearing words. The con side concludes the debate with further attack "Also a pointer for your future arguments on the same topic: learn to start sentence with capital letters, also if you want to nitpick about your spelling I can go on you 26 year old washed up community college dropout." In this example, the counting of swear words "stupid" would help with our prediction. More examples are below, which further illustrates my point. The first column is the number of times the pro debater says the word "stupid" and the second column is the number of time the con debater says the word "stupid". Almost all of the times when the pro side uses the word "stupid" more, they loses the debate.

	pro stupid count	con stupid count	title
603	19	24	The debate.org site rules should be more stric...
980	5	4	Legend of the Seeker stays true to The Sword o...
331	4	1	The 10 commandments are a JOKE!
340	4	1	god is not "one" therefore god is a fraud
1076	4	0	Tell me your thoughts on Abortion.
1118	4	0	Abortion
982	3	1	Term Lengths and Limitations
1207	3	1	macroevolution of humans has stopped for the f...
1223	3	1	Racist's are ignorant and are purely stupid.
984	2	6	Is Google Making Us Stupid?

(b) What lexicon did you chose to incorporate and why? How did you define features using this lexicon? Perform some data analysis to show that the features you extract would be useful for predicting the winner. Give two examples where this feature would help you identify the winner of a debate.

i. I decide to use the nrc_vad lexicon. Because the variables in this lexicon are continuous rather than categorical, I believe that it can provides a much richer set of information than the connotation lexicon. On the one hand, the connotation lexicon only has positive, negative, and neutral categories for each word. On the other hand, the vad lexicon has 3 real values that indicates the valence, arousal, and dominance of word. Particularly, the arousal aspect of a word would be helpful for predicting the winner presumably because the voters would be more likely to be persuaded by the debater with more rhetorical word usage that is correlated with the arousal metrics.

ii. For each side of a debate, I record the valence, arousal, and dominance of each word in the respective speeches. I then calculate the average valence, arousal, and dominance across all words in the debate.

iii. The following table presents the highest validation accuracy obtained by the respective vad and connotation lexicon across all other feature sets (i.e. trigrams, 2 linguistic features, etc). As you can see, the vad lexicon performs slightly better than the connotation lexicon. The difference, however, is marginal.

lexicon features	val_accuracy	train_accuracy
['vad']	0.769424	0.750628
['connotation']	0.746867	0.734925

iv. An example is the debate titled "Evidence for God". In the debate, the pro side has an average valence score of 0.009, arousal of 0.007, and dominance of 0.008. The con side has an average valence score of 0.008, arousal of 0.006, and dominance of 0.008. One of the example that the pro side gives in an argument is about rape: "in order for one to disagree with premise two, one must believe that an action like rape is just as "good" as an action like generosity, and that no objective distinction can be made between the nature of "goodness" of the two acts. " The pro side continues to use the example on rape several times in their subsequent arguments. Rapes are an especially heinous crime. The word is used rhetorically here to help the debater to establish its argument more firmly. The word "rape" also has a very high arousal of 0.8. Another example is a debate titled "The New and Old Testaments Should NOT Be Considered Moral Guides". In it, the pro debater uses the word rape 15 times to demonstrate an absolute moral truth, whereas the con side only uses it 2 times. The debate winner is the pro side.

(c) What user information do you think would be helpful in determining the winner of a debate and why? How will you incorporate this information as features into the model? Perform some data analysis to show that these features would be useful for predicting the winner of a debate. Give two examples where this feature would help you identify the winner of a debate.

   i. I believe that matching political ideology would have be helpful for predicting the winner of the debates. My hypothesis that voters are easily persuaded by views that they already agree with. So I believe that both religious and political ideology information are both useful for prediction. If I have to pick one, I believe that political ideology might be more useful. Even though there are 370 religious debates and 342 political debates in our training set, the rest of the categories such as society, education, health, economics, etc is more relevant to one's political ideology than religious ideology.

   ii. For each side of a debate, I plan to use a 3-dimension vector where each element represents the number of voters with different political ideologies that are not "Not Saying", the number of voters with "Not Saying" as their political ideologies, and the number voters with the same political ideologies as the debater. These numbers are then standard into percentages. Note that when the debater's political ideology is "Not Saying", all voters' political ideologies are treated as "Not Saying".

   iii. The following table shows the highest accuracy obtained on the validation set for all user features, which include alignment of political ideology, gender, education, relationship status, party association, and religious ideology, as well as cosine similarity between debaters and voters as defined in the paper "Exploring the Role of Prior Beliefs for Argument Persuasion". As you can see, political alignment and religious alignment features tend to perform the best with above 75% accuracy. Surprisingly, gender alignment and education alignment also plays a significant role.

   iv. For example, in the debate titled "The existence of free will as proposed by the bible", all voters have the same ideology as the pro debater as conservative while the con debater is "Not Saying". The result is that the pro debater is the winner. Similarly, in another debate titled "Medical Marijuana", all voters and the pro debaters are moderate and the con debater is "Other". The pro debater was voted as the winner.

Table 1: Ngram+lexicon+linguistic+users models

users features	val_accuracy
[political_align, gender_user_align]	0.786967
[religious_align, education_user_align]	0.766917
[religious_align, political_align]	0.766917
[political_align, party_user_align]	0.766917
[cosine_similarity , ethnicity_user_align]	0.764411
[political_align, ethnicity_user_align]	0.764411

(d) Did you find that different sets of features were useful for predicting the winner of a debate on religious topics vs. all other topics? Why do you think that's the case? What do your findings suggest about the factors that affect the outcome of a debate?

i. Yes I did find that different sets of features were useful for predicting the winner of a debate on religious topics vs. all other topics. Particularly, I found that speech length and the usage of personal pronouns of linguistic features, religious alignment of user features, and connotation of lexicon features are particularly effective for predicting winners of religious debates, yielding a 74% accuracy on the validation set. Their accuracy on non-religious debates is only 71%, which is quite low comparing to other features. On the other hand, features that work well with non-religious debates are the number of swear words and usage of personal pronouns, vad lexicon features, and the similarity of education level between the voters and the debaters. These features yield 76% on the non religious debates and 69% accuracy on the religious debates in terms of accuracy on the validation set.

4. Perpexity (10 points). You are given a training sequence of 100 animals that consists of 91 badgers and 1 each of 9 other animals (including a snake). You know that each non-badger occurs after ten badgers and the training sequence ends in a badger. Now, using add one smoothing, compute the bigram perplexity for each of the following test sequences where B denotes a badger and S denotes a snake. (Please ignore start-of-sequence and end-of-sequence characters).

Since every animal other than badgers must follow 10 badgers, we know that the sequence starts with a badger and that every continuous sequence of badgers must be of length 10. This leave 1 extra badger at the end. Hence, there are 9 continuous sequence of badgers each of length 10. Each one has 9 counts of the sequence(B, B). Hence, $C(B, B) = 81$. Since no non-badger animal is at the beginning or the end of the corpus, we know that $C(S, B) = C(B, S) = 1$ where S is a snake. For any other non-badger animals A_i, $C(S, A_i) = C(A_i, S) = C(S, S) = 0$ since a snake cannot follow non-badger animals. Under Laplace-smoothing, the following is true:

$$
\begin{aligned}
P(B|B) &= \frac{C(B,B)+1}{\sum_{i=1}^{8} C(B,A_i) + C(B,S) + C(B,B) + |V|} \\
&= \frac{81+1}{8+1+81+10} \\
&= \frac{82}{100} \\
P(S|B) &= \frac{C(B,S)+1}{\sum_{i=1}^{8} C(B,A_i) + C(B,S) + C(B,B) + |V|} \\
&= \frac{1+1}{8+1+81+10} \\
&= \frac{2}{100} \\
P(B|S) &= \frac{C(S,B)+1}{\sum_{i=1}^{8} C(S,A_i) + C(S,S) + C(S,B) + |V|} \\
&= \frac{1+1}{0+0+1+10} \\
&= \frac{2}{11} \\
P(S|S) &= \frac{C(S,S)+1}{\sum_{i=1}^{8} C(S,A_i) + C(S,S) + C(S,B) + |V|} \\
&= \frac{0+1}{0+0+1+10} \\
&= \frac{1}{11}.
\end{aligned}
$$

Let $n(B,B), n(B,S), n(S,B), n(S,S)$ be the number of occurences of the respective sequence of tokens in a new text. The perplexity of a new text W would be

$$
\begin{aligned}
PP(W) &= P(w_1 w_2 \dots w_N)^{\frac{-1}{N}} \\
&= (\Pi_{i=1}^{N} P(w_i|w_{i-1}))^{\frac{-1}{N}} \\
&= (P(B|B)^{n(B,B)} P(S|B)^{n(B,S)} P(B|S)^{n(S,B)} P(S|S)^{n(S,S)}))^{\frac{-1}{N}}.
\end{aligned}
$$

(a) B B B B B S B B B B

$$
\begin{aligned}
PP(BBBBBSBBBB) &= (P(B|B)^7 P(S|B)^1 P(B|S)^1))^{\frac{-1}{10}} \\
&= 2.0149
\end{aligned}
$$

(b) B B S S B B S S B B

$$
\begin{aligned}
PP(BBSSBBSSBB) &= (P(B|B)^3 P(S|B)^2 P(B|S)^2 P(S|S)^2))^{\frac{-1}{10}} \\
&= 5.2723
\end{aligned}
$$

(c) B B B B B B B B B B B B S

$$\begin{aligned} PP(BBBBBBBBBBBBS) &= (P(B|B)^{11} P(S|B)^1 P(B|S)^0 P(S|S)^0)^{\frac{-1}{13}} \\ &= 1.5981 \end{aligned}$$

5. Smoothing (10 points). Consider the following corpus of text: I went to the train station in San Francisco I went to the bus station in San Francisco I went to the metro station in San Francisco Francisco went to San Francisco for gas Suppose you build a bi-gram language model, with Katz backoff and absolute discounting with d = 0.75.

   (a) What is the most likely completion for the sentence below from following set of words station, Francisco, to? Calculate the probability of the full sequence. Show your work. I went to the gas __

      i. Let $\alpha(w_{i-1})$ be the remaining probability mass after absolute discount. Then

$$\alpha(w_{i-1}) = 1 - \sum_{v:C(w_{i-1},v)>0} \frac{c(w_{i-1}, v) - 0.75}{c(w_{i-1})}.$$

      In Katz backoff, we distribute the remaining probability proportional to the unigram probabilities. Specifically, if $C(w_{i-1}, w_i) = 0, P(w_i|w_{i-1}) \propto P(w_i)$. Summing it across all w_i that do not follow w_{i-1} gives $\sum_{v:C(w_{i-1},v)=0} P(v)$. Hence, if $C(w_{i-1}, w_i) = 0$,

$$P(w_i|w_{i-1}) = \alpha(w_{i-1}) \frac{P(w_i)}{\sum_{v:C(w_{i-1},v)=0} P(v)}.$$

      Otherwise,

$$P(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i) - 0.75}{c(w_{i-1})}.$$

ii.

$$
\begin{aligned}
P(I|BOS) &= \frac{C(BOS, I) - 0.75}{C(BOS)} \\
&= \frac{3 - 0.75}{4} \\
&= \frac{9}{16} \\
P(went|I) &= \frac{C(I, went) - 0.75}{C(I)} \\
&= \frac{3 - 0.75}{3} \\
&= \frac{12}{16} \\
P(to|went) &= \frac{C(went, to) - 0.75}{C(went)} \\
&= \frac{4 - 0.75}{4} \\
&= \frac{13}{16} \\
P(the|to) &= \frac{C(to, the) - 0.75}{C(to)} \\
&= \frac{3 - 0.75}{4} \\
&= \frac{9}{16}
\end{aligned}
$$

iii.

$$
\begin{aligned}
\sum_v C(v) &= 42 \\
\alpha(the) &= 1 - \frac{C(the, metro) - 0.75 + C(the, bus) - 0.75 + C(the, train) - 0.75}{C(the)} \\
&= 1 - \frac{1 + 1 + 1 - 3 \cdot 0.75}{3} \\
&= 0.75 \\
P(gas) &= P(metro) = P(train) = \frac{1}{42} \\
P(gas|the) &= \alpha(the) \frac{P(gas)}{\sum_{v:C(the,v)=0} P(v)} \\
&= 0.75 \frac{\frac{1}{42}}{1 - P(metro) - P(bus) - P(train)} \\
&= 0.75 \frac{\frac{1}{42}}{1 - \frac{1}{42} - \frac{1}{42} - \frac{1}{42}} \\
&= \frac{1}{52} \\
\mathbf{P(I, went, to, the, gas)} &= \frac{9 \cdot 12 \cdot 13 \cdot 9 \cdot 1}{16 \cdot 16 \cdot 16 \cdot 52} \\
&= \frac{243}{65536}
\end{aligned}
$$

iv.

$$
\begin{aligned}
\alpha(gas) &= 1 - \frac{\sum_v C(gas, v) - 0.75}{C(gas)} \\
&= 1 - \frac{C(gas, EOS) - 0.75}{C(gas)} \\
&= 1 - \frac{1 - 0.75}{1} \\
&= 0.75 \\
P(station) &= \frac{1}{42} \\
\sum_{v:C(gas,v)=0} P(v) &= 1 - P(EOS) = 1 - \frac{4}{42} = \frac{38}{42} \\
P(station|gas) &= \alpha(gas)\frac{P(station)}{\sum_{v:C(gas,v)=0} P(v)} \\
&= 0.75 \cdot \frac{1}{42} \cdot \frac{42}{38} \\
&= \frac{3}{152} \\
\mathbf{P(I, went, to, the, gas, station)} &= \frac{243}{65536} \cdot \frac{3}{152} \\
&= 0.00007318
\end{aligned}
$$

v.

$$
\begin{aligned}
P(francisco|gas) &= \alpha(gas)\frac{P(francisco)}{\sum_{v:C(gas,v)=0} P(v)} \\
&= 0.75 \cdot \frac{5}{42} \cdot \frac{42}{38} \\
&= \frac{15}{152} \\
\mathbf{P(I, went, to, the, gas, francisco)} &= \frac{243}{65536} \cdot \frac{15}{152} \\
&= 0.0003659
\end{aligned}
$$

vi.

$$
\begin{aligned}
P(to|gas) &= \alpha(gas)\frac{P(to)}{\sum_{v:C(gas,v)=0} P(v)} \\
&= 0.75 \cdot \frac{4}{42} \cdot \frac{42}{38} \\
&= \frac{12}{152} \\
\mathbf{P(I, went, to, the, gas, to)} &= \frac{243}{65536} \cdot \frac{12}{152} \\
&= 0.0002927
\end{aligned}
$$

vii. The most likely completion is "francisco".

(b) Is what you get above consistent with what you would expect to be the most probable completion? If not, can you suggest a different smoothing method that could fix this issue?

i. This is inconsistent with what I expect to be. It only follows after "san". Since "francisco" and "station" both do not follow "gas", its conditional probability after "gas" is proportioned to its unigram counts in the corpus. Because "francisco" appears more frequently in the text, it is attributed with higher probability. One way to fix this is to take into consideration of how likely a word is to be after other words. For example, "station" is a much flexible word. It follows after "train", "bus", and "metro". In contrast, the word "francisco" is a inflexible. Instead of using counts, we can adjust P("station") by the number of bi-grams in which "station" follows after $|\{v \mid C(v, station) > 0\}|$. If we sum up the adjustment term across all words that do not follow "gas", we get

$$\begin{aligned}\sum_{w^*:C(gas,w^*)=0} |\{v \mid C(v, w^*) > 0\}| &= \sum_{w^*:C(gas,w^*)=0} |\{v \mid C(v, w^*) > 0\}| \\ &= |\{v \mid C(gas, v) > 0\}|.\end{aligned}$$

Hence, in general if $C(w_{i-1}, w_i) = 0$,

$$\alpha(w_{i-1})\frac{|\{v \mid C(v, w_i) > 0\}|}{|\{v \mid C(w_{i-1}, v) > 0\}|}.$$

Additionally, we can also put in weights to determine whether we should look at the unigram count more or the flexibility more when distributing the remaining probability $\alpha(w_{i-1})$.

$$\alpha(w_{i-1})(\lambda_2\frac{|\{v \mid C(v, w_i) > 0\}|}{|\{v \mid C(w_{i-1}, v) > 0\}|} + \lambda_1\frac{P(w_i)}{\sum_{v:C(w_{i-1},v)=0} P(v)}).$$

Table 2: Majority

val_accuracy	train_accuracy
0.528822	0.575377

Table 3: Ngram

ngram features	val_accuracy	train_accuracy
['trigram']	0.726817	0.888191

Table 4: Ngram+lexicon models

lexicon features	val_accuracy	train_accuracy
['vad']	0.734336	0.807789
['connotation']	0.729323	0.726759

Table 5: Ngram+lexicon+linguistic models

	lexicon features	linguistic features	val_accuracy	train_accuracy
44	['vad']	['reference_to_opponent', 'swear_words']	0.769424	0.750628
51	['vad']	['swear_words', 'personal_pronouns']	0.764411	0.754397
55	['vad']	['swear_words', 'number']	0.754386	0.742462
52	['vad']	['swear_words', 'questions']	0.754386	0.736809
47	['vad']	['reference_to_opponent', 'websites']	0.751880	0.750628
37	['vad']	['length', 'swear_words']	0.751880	0.744975
56	['vad']	['swear_words', 'modal_verb']	0.751880	0.744347
60	['vad']	['personal_pronouns', 'number']	0.749373	0.744975
66	['vad']	['websites', 'exclamation']	0.749373	0.744347
36	['vad']	['length', 'reference_to_opponent']	0.749373	0.748744
45	['vad']	['reference_to_opponent', 'personal_pronouns']	0.746867	0.745603
49	['vad']	['reference_to_opponent', 'number']	0.746867	0.744975
57	['vad']	['personal_pronouns', 'questions']	0.746867	0.750000
59	['vad']	['personal_pronouns', 'exclamation']	0.746867	0.736809
2	['connotation']	['length', 'personal_pronouns']	0.746867	0.734925
67	['vad']	['websites', 'number']	0.744361	0.744347
53	['vad']	['swear_words', 'websites']	0.741855	0.748116
29	['connotation']	['questions', 'modal_verb']	0.741855	0.728015
62	['vad']	['questions', 'websites']	0.741855	0.743090
14	['connotation']	['reference_to_opponent', 'modal_verb']	0.741855	0.736809
42	['vad']	['length', 'number']	0.741855	0.744347
46	['vad']	['reference_to_opponent', 'questions']	0.741855	0.742462
68	['vad']	['websites', 'modal_verb']	0.741855	0.746859
40	['vad']	['length', 'websites']	0.736842	0.748116
12	['connotation']	['reference_to_opponent', 'exclamation']	0.736842	0.733040
61	['vad']	['personal_pronouns', 'modal_verb']	0.734336	0.750628
58	['vad']	['personal_pronouns', 'websites']	0.734336	0.757538
43	['vad']	['length', 'modal_verb']	0.731830	0.741206
64	['vad']	['questions', 'number']	0.729323	0.734925
18	['connotation']	['swear_words', 'exclamation']	0.729323	0.724874
35	['connotation']	['number', 'modal_verb']	0.729323	0.726131
39	['vad']	['length', 'questions']	0.729323	0.734296
41	['vad']	['length', 'exclamation']	0.729323	0.744347
33	['connotation']	['exclamation', 'number']	0.726817	0.724246
27	['connotation']	['questions', 'exclamation']	0.726817	0.724874
34	['connotation']	['exclamation', 'modal_verb']	0.721805	0.721734

Table 6: Ngram+lexicon+linguistic+users models
with vad lexicon feature, reference to opponent counts and swear words as linguistic features

users features	val_accuracy
[political_align, gender_user_align]	0.786967
[religious_align, education_user_align]	0.766917
[religious_align, political_align]	0.766917
[political_align, party_user_align]	0.766917
[cosine_similarity , ethnicity_user_align]	0.764411
[political_align, ethnicity_user_align]	0.764411
[political_align, education_user_align]	0.764411
[cosine_similarity , relationship_user_align]	0.761905
[education_user_align, ethnicity_user_align]	0.761905
[religious_align, relationship_user_align]	0.759398
[education_user_align, party_user_align]	0.759398
[religious_align, ethnicity_user_align]	0.756892
[cosine_similarity , political_align]	0.756892
[political_align, relationship_user_align]	0.756892
[education_user_align, relationship_user_align]	0.751880
[religious_align, party_user_align]	0.749373
[gender_user_align, relationship_user_align]	0.744361
[religious_align, gender_user_align]	0.744361
[party_user_align, relationship_user_align]	0.744361
[cosine_similarity , religious_align]	0.744361
[cosine_similarity , party_user_align]	0.744361
[education_user_align, gender_user_align]	0.741855
[ethnicity_user_align, relationship_user_align]	0.741855
[cosine_similarity , education_user_align]	0.739348
[gender_user_align, ethnicity_user_align]	0.739348
[cosine_similarity , gender_user_align]	0.739348
[party_user_align, ethnicity_user_align]	0.736842
[party_user_align, gender_user_align]	0.734336

Table 7: Ngram+lexicon+linguistic+users models
separate training between religious and non-religious debates

religious validation accuracy	non religious validation accuracy	validation accuracy
0.751880	0.741935	0.744253
0.761905	0.731183	0.738344
0.759398	0.731183	0.737759
0.719298	0.741935	0.736659
0.746867	0.731183	0.734839
0.734336	0.731183	0.731918
0.726817	0.731183	0.730165
0.761905	0.720430	0.730097
0.749373	0.709677	0.718930
0.736842	0.709677	0.716009
0.736842	0.709677	0.716009
0.764411	0.698925	0.714188
0.746867	0.698925	0.710099
0.744361	0.698925	0.709515
0.731830	0.698925	0.706594
0.764411	0.688172	0.705942
0.746867	0.688172	0.701853
0.741855	0.688172	0.700685
0.736842	0.688172	0.699516
0.734336	0.688172	0.698932
0.729323	0.688172	0.697764
0.754386	0.677419	0.695359
0.751880	0.677419	0.694775
0.749373	0.677419	0.694191
0.749373	0.677419	0.694191
0.744361	0.677419	0.693022
0.736842	0.677419	0.691270
0.721805	0.677419	0.687765
0.744361	0.666667	0.684776
0.741855	0.666667	0.684192
0.739348	0.666667	0.683608
0.739348	0.666667	0.683608
0.729323	0.666667	0.681271
0.759398	0.655914	0.680034
0.741855	0.655914	0.675945
0.736842	0.655914	0.674777