1) briefly state the goal of the project

incorporate implicit discourse relation disambiguation for sentiment analysis

importance of identifying context change (polarity)

tagger? Given previous polarity and discourse relation

unsupervised tagger but supervised polarity score

  2) briefly state the research hypotheses you will test

  (what question(s) do you want to answer?)

  3) briefly state what you plan to do

  4) discuss in detail the previous research on this topic and how it related to

  your planned work

Word pairs for sentiment

Use imdb corpus?

Discuss other work on discourse and sentiment

Adjective pairs (hatsivassiloglou)

Graphical models for discourse and sentiment

Other work looking at twitter

This work would try to also identify implicit discourse relations

Using word pairs

Use word2vec model

Section on PDTB

Explain explicit discourse markers and discourse relations

Explicit discourse connectives can help to identify their associated discourse class with 95% accuracy (cite pitler)

Don’t need fine-grained discourse relations for this project

Implicit are much more difficult

At least 1/3 of comparison relations are implicit so if there is a polarity switch we could not detect it from the discourse markers alone

Section on Word Pairs

Cite marcu, pitler, nenkova, biran

Word pairs for sentiment

Analyze positive and negative sentiment for each discourse marker

Also do neutral?

+/+, -/-, +/-, -/+, 0/0, 0/+, 0/-, +/0, -/0

How to determine sentiment?

Maybe compare several methods

Count positive and negative words?

Use model for sentiment?

Use net difference of positive and negative words?

Opinion Finder

Distant supervision?

Don’t really have data that has clauses annotated for sentiment

There is some data that has individual words marked

ONLY train on data where discourse marker is present?

what corpus to use? Compare different corpora?

Gigaword?

Twitter?

Wikipedia?

Movie reviews?

Train word pairs on separate corpus

Problem is domain specific (words in movies vs elsewhere)

How to train model? Can use discourse markers but need explicit marker for sentiment, or annotated corpus

word pairs with matrix decomposition?

Finally, use these word pair features to train a classifier on some other corpus

Use amazon product review

Train word pair model on some set aside data

Create two scores: one as if the review was positive and one for negative

Use these scores as features during training

Use opinionfinder to determine changes in polarity

+/+, -/-, +/-, -/+, 0/0, 0/+, 0/-, +/0, -/0

Weight the cosine similarity scores somehow using this

Use word pairs directly

Compare with aggregated word pairs

Also consider within-sentence discourse relations

May shift within a sentence

Features:

Discourse connective appears within first N words

For word pairs

Try word pairs as features directly

May run into high dimensionality issue

Try aggregated word pairs using all discourse markers

(still very sparse, so…)

try using word embeddings to learn a model

PU learning?

Try aggregated word pairs using only contrastive/non-contrastive markers

Manual clustering of word pairs

Automated clustering

Kmeans

K=2 (contrast or not)

K=3 (contrast or not or neutral)

Non-parametric

Constrained

But/and not in same class

But/however in same class?

Also do this for words at beginning of sentence to see if we can learn new unigram connectives

Use annotated subjectivity data to generate subjectivity transitions

Learn which words or phrases are likely to start a transition

Learn word pairs for transitions and non-transitions

Use opinion finder to learn these

Use twitter data to create separate models for positive and negative data

204 models total, try to cluster these

Also try on out of domain

Create word pairs from twitter data

Coherence model for change of subject

Structural features

Sentence length

Global features

(Relative?) Position of sentence

Slope/incoming/outgoing in spline

Read sentiment subgroup paper from 4/16

TODO:

Debug use of hidden vars? DONE

Use Stanford Treebank instead of opinion finder for hidden vars? NO TIME

Chi squared test comparing positive and negative word pairs for each marker

(take top K for polarity features)

Try difference rather than absolute (does this matter? Still linear combo)

(n-1) degrees of freedom

Also do chi-squared test on sentiment scored pairs (normalized?)

(n-1)\*(n-1) degrees of freedom

Use stanford treebank scores and paired scores as features

Chi squared feature selection

Combine markers into one model

Constrained clustering

Combine pairs that score highly

rebuild word pairs for entire tweet corpus

take difference between subjective (twitter) and non-subjective (news) as a subjectivity feature (not difference, just use both) - DONE

word pair features

all discourse models - DONE

top K discourse models according to p value

full sentiment model

same top 2 but for tf-idf word pairs

same top 2 but for weight vectors?

Structural features (subjectivity) - DONE

Sentence length

Position in document (x/total sentences)

Global features (subjectivity)

For whissell and/or LIWC categories

If it is a global minimum

If it is a global maximum

Difference between t and t-1

Difference between t and t+1

[1] Running nohup ~/PDTB/scikit/bin/python3.2 -u generateSLEFormat.py marker\_cos\_model\_all 10000 --polarityFeatures discourse\_model --iterations 10 > marker\_cos\_model\_all\_log &

[2] Running nohup ~/PDTB/scikit/bin/python3.2 generateSLEFormat.py baseline\_all 10000 --polarityFeatures '' --subjectivityFeatures '' --iterations 10 > baseline\_all\_log &

[3]- Running nohup ~/PDTB/scikit/bin/python3.2 generateSLEFormat.py marker\_all 10000 --polarityFeatures '' --subjectivityFeatures markers --iterations 10 > marker\_all\_log &

[4]+ Running nohup ~/PDTB/scikit/bin/python3.2 generateSLEFormat.py marker\_cos\_all 10000 --iterations 10 > marker\_cos\_all\_log &

include all sentiment model

do only top k models

And/or of discourse and all?

add as document features?

Have each sentence vote?

re-run structural

re-run obj

rerun pcos

Problems:

What to do if no subjective sentence?

Hyperparameters?

No data set or code available

Opinion finder not accurate?

Future work:

Data sparsity

Clustering

Word embeddings

Why not language model?

Discourse indicates change of state

Investigate other markers that could indicate change of state

Use addDiscourse to identify actual discourse markers

SemEval 2013 data?

baseline\_all\_log:**Zero**/one-error on test set: 12.84% (4358 correct, 642 incorrect, 5000 total)

baseline\_all\_log:**Zero**/one-error on test set: 13.24% (21687 correct, 3309 incorrect, 24996 total)

marker\_all\_log:**Zero**/one-error on test set: 12.52% (4374 correct, 626 incorrect, 5000 total)

marker\_all\_log:**Zero**/one-error on test set: 12.98% (21752 correct, 3244 incorrect, 24996 total)

marker\_cos\_all\_log:**Zero**/one-error on test set: 12.48% (4376 correct, 624 incorrect, 5000 total)

marker\_cos\_all\_log:**Zero**/one-error on test set: 13.21% (21693 correct, 3303 incorrect, 24996 total)

marker\_cos\_model\_all\_log:**Zero**/one-error on test set: 12.42% (4379 correct, 621 incorrect, 5000 total)

marker\_cos\_model\_all\_log:**Zero**/one-error on test set: 13.43% (21640 correct, 3356 incorrect, 24996 total)

(not done)

marker\_cos\_model\_obj\_all\_log

Zero/one-error on test set: 12.52% (4374 correct, 626 incorrect, **5000** total)

Zero/one-error on test set: 13.25% (21685 correct, 3311 incorrect, **24996** total)

marker\_cos\_model\_topk\_all\_log

Zero/one-error on test set: 12.24% (4388 correct, 612 incorrect, **5000** total)

Zero/one-error on test set: 12.94% (21761 correct, 3235 incorrect, **24996** total)

baseline/c1000.l50/svm\_classify\_log:**Zero**/one-error on test set: 12.38% (4381 correct, 619 incorrect, 5000 total)

baseline/c1000.l50/test\_log:**Zero**/one-error on test set: 13.19% (21698 correct, 3298 incorrect, 24996 total)

marker\_only\_debug\_log:**Zero**/one-error on test set: 12.74% (4363 correct, 637 incorrect, 5000 total)

marker\_only\_debug\_log:Zero/one-error on test set: 13.11% (21719 correct, 3277 incorrect, **24996** total)

87.2

-

82%

90%

87.5%

84%

92.5%

87.18

-

80.6

91.25

87

84

93

87.24

-

82.5

90.6

86.25

85

91.88

86.58

-

80

89.4

87.5

84

92

baseline

12.50% (4375 correct, 625 incorrect, 5000 total)

13.19% (21698 correct, 3298 incorrect, 24996 total)

cos\_tmp2\_pcos2\_log

12.60

13.26

marker\_only\_log

12.32 (4384 correct, 616 incorrect, 5000 total)

13.21 (21695 correct, 3301 incorrect, 24996 total)

marker\_only\_debug\_log

13.11% (21719 correct, 3277 incorrect, 24996 total)

marker\_and\_cos\_debug\_log

12.20

13.01

marker\_and\_cos\_obj\_debug\_log (debug? Rerun this)

12.20 (4390 correct, 610 incorrect, 5000 total)

13.01 (21745 correct, 3251 incorrect, 24996 total)

change to do hillclimb???

Hypothesis

Discourse aware modeling of sentiment can help

1. within sentence discourse relations
2. implicit across sentence relations

Related work

1. lightweight twitter
2. latent SVM using discourse markers
3. sentiment using recurrent neural nets

much slower on same data

requires parsing and hand-labeled data

domain-specific

Data

Sentiment140

IMDB corpus (maas et al)

Part 1 Results (1-2 slides)

Lightweight modeling of discourse

Compare unigrams vs word pairs, ordered by improvement in p-value

Part 2