Hector Cuevas
January 8, 2011

Agreement, Disagreement & Style Matching

Agreement and Disagreement

- *Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies*
- M. Galley, K. McKeown, J. Hirschberg, and E. Shriberg. 2004

Motivation

- Automated summarization of multi-participant meetings
- Identifying agreement and disagreement between participants will help

Previous Work

- Hillard et al, 2003
- Feasible using textual, durational, and acoustic features
- Build on this, improve when contextual information taken into account

Approach

- Identify adjacency pairs using maximum entropy ranking
  - Based on set of lexical, durational, and structural features that look both forward and backward in the discourse
- Acquire, and process, knowledge about who speaks to whom

Corpus

- ICSI Meeting Corpus
  - 75 meetings collected at the International Computer Science Institute (ICSI)
  - Naturally occurring, regular weekly meetings
  - Meetings just under 1 hour each
  - Average of 6.5 participants
- Labeled with adjacency pairs
  - Provide information about speaker interaction
- Structure of conversations as paired utterances

Corpus

- 7 meetings segmented into spurts
  - Periods of speech with no pauses greater than .5 seconds
- Used spurt segmentation instead of sentence segmentation
  - Easy to obtain and automate the system
- Labeled with 4 categories
  - Agreement, disagreement, backchannel, and other

Adjacency Pairs

- Given the second element (B) of an adjacency pair, determine who is the speaker of the first element (A)
- Baseline: select speaker before the occurrence of utterance B.
  - Selects the correct speaker 79.8% of cases in 50 meetings annotated with adjacency pairs

Maximum Entropy Ranking

- Four categories of features
  - Structural, durational, lexical, and dialog act information
- Structural
  - Ordering and overlap of spurts information
- Lexical
  - Count based – remove infrequent and uninformative words

Maximum Entropy Results

Structural features:

- number of speakers taking the floor between A and B
- number of spurts between A and B
- number of spurts of speaker B between A and B
- do A and B overlap?

Durational features:

- duration of A
- if A and B do not overlap: time separating A and B
- if they do overlap: duration of overlap
- seconds of overlap with any other speaker
- speech rate in A

Lexical features:

- number of words in A
- number of content words in A
- ratio of words of A (respectively B) that are also in B (respectively A)
- ratio of content words of A (respectively B) that are also in B (respectively A)
- number of n-grams present both in A and B (we built 3 features for n ranging from 2 to 4)
- first and last word of A
- number of instances at any position of A of each cue word listed in (Hirschberg and Litman, 1994)
- does A contain the first/last name of speaker B?

Table 1. Speaker ranking features

- DA features improve
  - Not included, difficult to acquire automatically

Feature sets	Accuracy
Baseline	79.80%
Structural	83.97%
Durational	84.71%
Lexical	75.43%
Structural and durational	87.88%
All	89.38%
All (only backward looking)	86.99%
All (Gaussian smoothing, FS)	90.20%

Table 2. Speaker ranking accuracy

Agreement/Disagreement Classification

Structural features:

- is the previous/next spurt of the same speaker?
- is the previous/next spurt involving the same B speaker?

Durational features:

- duration of the spurt
- seconds of overlap with any other speaker
- seconds of silence during the spurt
- speech rate in the spurt

Lexical features:

- number of words in the spurt
- number of content words in the spurt
- perplexity of the spurt with respect to four language models, one for each class
- first and last word of the spurt
- number of instances of adjectives with positive polarity (Hatzivassiloglou and McKeown, 1997)
- idem, with adjectives of negative polarity
- number of instances in the spurt of each cue phrase and agreement/disagreement token listed in (Hirschberg and Litman, 1994; Cohen, 2002)

Table 3. Local features for agreement and disagreement classification

- Features found most helpful at identifying agreements and disagreements
- Duration

Agreement/Disagreement Assumptions

- Previous tag dependency
  - Tag is influenced by its predecessor
- Same-interactants previous tag dependency
  - If B disagrees with A, B is likely to disagree with A in his or her next speech addressing A.
- Reflexivity
  - Speaker B influenced by what A said last to B.
- Transitivity
  - Speaker X agrees with A, then B disagrees with X. → B in disagreement with A.

Testing Assumptions

- Long range dependencies undesirable
- Made Markov assumption limiting context to arbitrarily chosen value *N* (*N* = 10)

	$p(c_i \mid c_{i-1})$	$p(c_i^{B\to A} \mid \mathrm{pred}_{B\to A}(c_i^{B\to A}))$	$p(c_i^{B\to A} \mid \mathrm{pred}_{A\to B}(c_i^{B\to A}))$
$p(\textsc{Agree} \mid \textsc{Agree})$	**.213**	**.250**	.175
$p(\textsc{Other} \mid \textsc{Agree})$	.713	.643	.737
$p(\textsc{Disagree} \mid \textsc{Agree})$	.073	.107	.088
$p(\textsc{Agree} \mid \textsc{Other})$	.187	.115	.177
$p(\textsc{Other} \mid \textsc{Other})$	.714	.784	.710
$p(\textsc{Disagree} \mid \textsc{Other})$	.098	.100	.113
$p(\textsc{Agree} \mid \textsc{Disagree})$	.139	.087	**.234**
$p(\textsc{Other} \mid \textsc{Disagree})$	.651	.652	.638
$p(\textsc{Disagree} \mid \textsc{Disagree})$	**.209**	**.261**	.128

Table 4. Contextual dependencies (previous tag, same-interactants previous tag, and reflexivity)

Testing Assumptions

- Previous tag dependency (A/D)
  - 18.8%/10.6%
  - 13.9%/20.9% preceded by DISAGREE
  - 21.3%/7.3% preceded by AGREE
- Same-interactants previous tag dependency
  - 26.1% B disagrees with A, continue to disagree in next exchange
  - 25% for agree

Testing Assumptions

- Reflexivity (P(AGREE) = 0.188)
  - P(AGREE | AGREE) = .175
  - P(AGREE | DISAGREE) = .234
- Transitivity
  - X agrees with A and B, B agrees with A 22.5%
  - Cannot conclude disagreement with a disagreement is equivalent to agreement
    - May not concern the same propositional content

Sequence Classification

- HMM maximize joint likelihood of training data
  - Enumerate all possible sequences of observations
- Conditional Markov Models (CMM) address concern
  - Interacting features and long range dependencies
- Bayesian network model
  - Incorporate more than one label dependency
  - Take 4 pragmatic contextual dependencies into account
- Compute most probable sequence using a left-to-right decoding using beam search (N = 100)

Setup

- 8135 spurts for training and testing
- Recreated setup of Hillard et al, 2003
  - 3-way classification
- N-fold cross-validation in four-way classification task
- Hand-labeled data, single meeting, for testing and the rest for training

Results

- With 3 local feature sets only, obtain better results than Hillard et al, 2003
  - Additional features not exploited previously
    - Structural features & adjective polarity
  - Learning algorithm may be more accurate
- Corroborates findings that lexical information most helpful local features
- Incorporate label-dependency features (pragmatic influences) improve performance about 1%

Results

- Identifying adjacency pairs can help
- Has applications to multi-party dialog act classification

Feature sets	Accuracy
(Hillard et al., 2003)	82%
Lexical	84.95%
Structural and durational	71.23%
All (no label dependencies)	85.62%
All (with label dependencies)	86.92%

Table 6. 3-way classification accuracy

Feature sets	Label dep.	No label dep.
Lexical	83.54%	82.62%
Structural, durational	62.10%	58.86%
All	84.07%	83.11%

Table 7. 4-way classification accuracy

Style Matching

- *Language Style Matching as a Predictor of Social Dynamics in Small Groups*
- A. Gonzales, J. Hancock, and J. Pennebaker. 2010

Motivation

- Mimicry frequently linked to social processes
- People who like one another produce similar styles of speech
- Mimicry also linked to team performance
- May facilitate language production and comprehension

Previous Work

- Nonverbal mimicry by Condon and Ogston (1966)
- Replayed videotape of social interactions
  - Each change in movement recorded and coded

Approach

- Linguistic style matching (LSM)
  - Automated text analysis – parse conversations into psychologically relevant dimensions
  - Function words – have proven useful in identifying relationships between language and social psychological states (Campbell & Pennebaker, 2003)
  - Measures the degree of similar rates of function words in two or more people's dialogue

LSM and Social Dynamics

- Investigated LSM ability to predict cohesiveness and task performance
  - Two well-known aspects of social dynamics in small groups
- Cohesiveness
  - Social synchrony metric of positively functioning social dynamics
- Task Performance
  - Improved communication synchrony improves group's performance

Communication Media

- Will verbal mimicry occur at equal rates in face-to-face (FTF) and online groups?
- Can LSM predict cohesiveness and task performance in both media?

Language Indicators

- Are these indicators of cohesiveness and task performance?
- Mimicry
- Simple verbal features of interaction
  - Word count
  - Pronoun patterns
  - Future and achievement-oriented language

Testing Overview

- Work groups of 4-6 work together on a problem-solving task (same sex)
- Half in same room (FTF)
- Remaining using online chat technology
- 35 minutes
  - 10 mins – get to know one another
  - 25 mins – come up with answers to 22 different questions
- Interaction Rating Questionnaire (IRQ)

Overview

- Language transcribed compared with outcomes of group cohesiveness and performance
  - Linguistic Inquiry and Word Count (LIWC)

LSM scores – Group Cohesiveness

- Degree to which each group member used 9 types of function words
  - Auxilary verbs, articles, common adverbs, personal pronouns, indefinite pronouns, prepositions, negations, conjunctions, and quantifiers
- Each person's language with overall percentage of remaining group members

$$pp1 = 1 - (|pp1 - ppG|/(pp1 + ppG)),$$
$$pp2 = 1 - (|pp2 - ppG|/(pp2 + ppG)),$$
$$pp3 = 1 - (|pp3 - ppG|/(pp3 + ppG)),$$
$$pp4 = 1 - (|pp4 - ppG|/(pp4 + ppG)),$$

$$\text{Group ppLSM} = (pp1 + pp2 + pp3 + pp4)/4,$$

LSM Scores – Group Cohesiveness

- Verbal matching significant indicator of how well group members like one another
- Also examined communication medium
  - LSM cohesion prediction not affected by medium

Table 1. Mean Category LSM Score and Total LSM Score

Category LSM Score	Examples	*M*	*SD*
Adverb	completely, often	0.89	0.06
Article	a, an, the	0.88	0.06
Auxiliary verb	am, have	0.91	0.04
Conjunction	and, but, or	0.85	0.06
Indefinite pronoun	it, those	0.90	0.04
Negation	no, not, never	0.79	0.10
Personal pronoun	I, you, we	0.92	0.04
Preposition	at, for, into	0.91	0.05
Quantifier	all, few, some	0.85	0.07
Total LSM score		0.88	0.03

Note: Mean scores refer to the average LSM score between each group member and the sum of the remaining group members. *SD* refers to the standard deviation of mean LSM scores across all 75 groups. Total LSM score is calculated by averaging the category LSM scores. LSM = language style matching.

LSM Scores – Task Performance

- Determined by percentage correct responses to 22 short questions
- FTF produced significant positive relationship between performance and LSM
- Online groups negative relationship between LSM and performance
- Group size improved group performance

Additional Language Indicators

- Word count positive predictor of group cohesiveness
- Groups using fewer first-person plural pronouns were more cohesive
- Positive relationship between use of future-oriented language and task performance
- Achievement-oriented language negatively related to task performance

Discussion

- LSM metric predicted group cohesiveness regardless of:
  - Communication medium
  - Gender of the group
  - How many individuals in the group
- Positive relationship between LSM metric and group cohesiveness

Discussion

- Positive relationship between LSM metric and task performance
  - Only in the FTF condition

LSM Metric

- Automated uniquely efficient tool
- Function words effective measures of mimicry
  - Frequently used, unconsciously produced, and context independent
- Mimicry occurs in both online and offline groups.

Limitations

- Shifting attention between the computer and the almanac detracted from interacting verbally in group

Questions