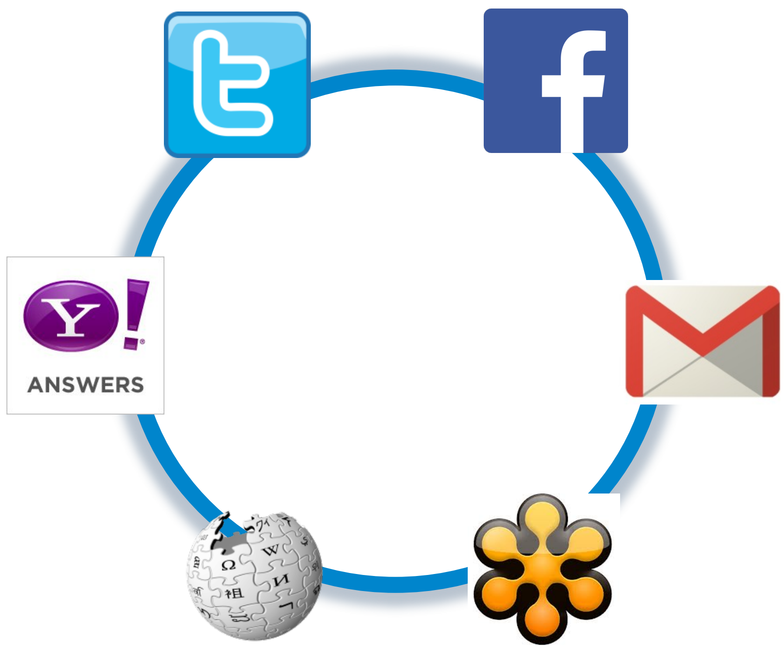
A Piece of My Mind: A Sentiment Analysis Approach for Online Dispute Detection

Lu Wang and Claire Cardie
Department of Computer Science
Cornell University

Cornell University

Introduction



- According to Wikipedia, there are more than 4.5M articles in English Wikipedia alone.
- About 21.6M users.
- About 130K registered editors.

A Disputed Discussion

Emy: I think everyone is forgetting that my previous image was the lead image for well over a year! … *Massimo is the one who began the "edit war"*...

Massimo: *I'm not going to start a debate about who started the fight, since it is childish and pointless*... As for your new image... I'm sorry to say so, but it is grossly over processed...

Emy: Yes, your camera has slightly higher resolution than mine. *I'm glad you paid more money for a camera than I did. Congrats. I appreciate your constructive criticism. Thank you.*

Massimo: First of all, I want to make clear that this is not personal. I just want to have the best picture as a lead for the article.

Emy: *Wow, I am really enjoying this photography debate. It is seriously making my work day so much more enjoyable!* … don't make assumptions you know nothing about. *Really, grow up. … Sound good?*

Massimo: *I do feel it is a pity, that you turned out to be a sore loser.*

The Problem: Online Dispute Detection

"Sometimes I think the collaborative process would work better without you."

[Credit: https://www.cartoonbank.com]

The Problem: Online Dispute Detection

- Facilitate collaboration

[Credit: http://wondermark.com]

- Identify controversial topics
- Analyze user relations
- Predict stance

Our Objectives

- Detecting the online disputes automatically
- Predicting disputes on a newly constructed dataset of scale.
- Understanding whether linguistic features, e.g. sentiment flow, are importance for dispute detection.

Previous Work

- Analyzed dispute-laden content to discover features correlated with conflicts and disputes
  - Kittur et al. (2007): edit history
  - Billings and Watts (2010): dispute resolution
  - Yasseri et al. (2012): temporal characteristics
  - Kraut and Resnick (2012): design of successful online communities
- However, they all rely on small number of manually selected discussions known to involve disputes.

Roadmap

- A dispute corpus constructed from Wikipedia
- Online dispute detection
  - Sentence-level sentiment prediction
  - Dispute detection
- Conclusion

Roadmap

- A dispute corpus constructed from Wikipedia
- Online dispute detection
  - Sentence-level sentiment prediction
  - Dispute detection
- Conclusion

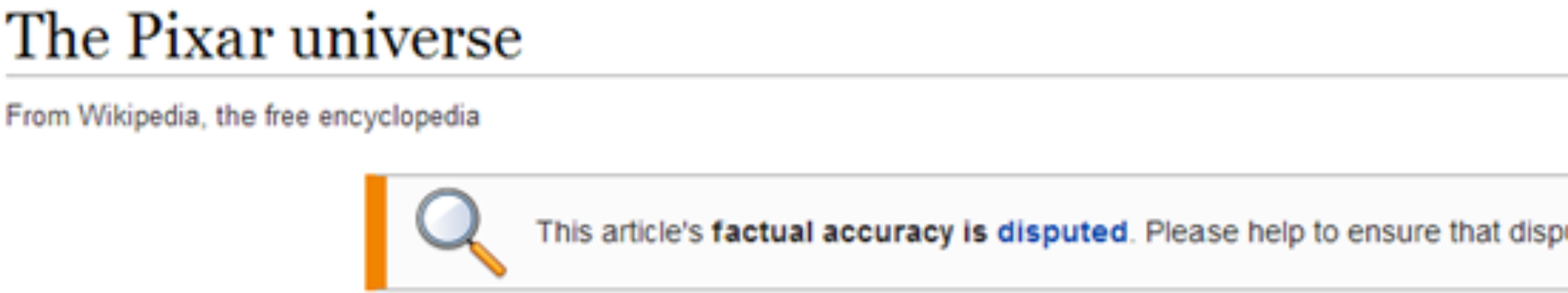
A Dispute Corpus Constructed from Wikipedia

The Pixar universe

From Wikipedia, the free encyclopedia

This article's **factual accuracy is disputed**. Please help to ensure that disp

A Dispute Corpus Constructed from Wikipedia



- *Disputed*
- *TotallyDisputed*
- *DisputedSection*
- *TotallyDisputedSection*
- *POV*

- 2013-03-04 Wikipedia data dump
- Result in 19,071 talk pages

A Dispute Corpus Constructed from Wikipedia

- **Step 1: Get Talk Pages of Disputed Articles**

- **Step 2: Get Discussions with Disputes.**
  - 3609 discussions are collected

- **Step 3: Get Discussions without Disputes.**
  - 3609 non-dispute discussions are randomly selected.
  - We consider non-dispute discussions with at least 3 distinct speakers and 10 turns.
  - The average turn numbers for dispute and non-dispute discussions are 45.03 and 22.95, respectively

Roadmap

- A dispute corpus constructed from Wikipedia
- Online dispute detection
  - Sentence-level sentiment prediction
  - Dispute detection
- Conclusion

Sentence-Level Sentiment Prediction

- Input: sentences $x=\{x_1, ..., x_n\}$ from a single turn
- Output: sequence of sentiment labels $y=\{y_1, ..., y_n\}$, where $y_i \in \{NN, N, O, P, PP\}$

- NN: very negative
- N: negative
- O: neutral
- P: positive
- PP: very positive

- Partial order: $NN \leq N \leq O \leq P \leq PP$

Sentence-Level Sentiment Prediction

- **Isotonic Conditional Random Fields (CRF)**
  - Mao and Lebanon (2007) proposed isotonic CRF to predict sentiment in movie reviews.

  - Encode domain knowledge through isotonic constraints on model parameters.

Isotonic CRF

$$p(y|x) = \frac{1}{Z(x)} \exp\left(\sum_{i} \sum_{\sigma,\tau} \lambda_{<\sigma,\tau>} f_{<\sigma,\tau>}(y_{i-1}, y_i) + \sum_{i} \sum_{\sigma,w} \mu_{<\sigma,w>} g_{<\sigma,w>}(y_{i-1}, x_i)\right)$$

- $f_{<\sigma,\tau>}$, $g_{<\sigma,w>}$ are feature functions, $\lambda_{<\sigma,\tau>}$, $\mu_{<\sigma,w>}$ are the parameters when y_{i-1}, y_i, x_i take values of λ, τ, w.
- Lexicon $M = M_p \cup M_n$, where M_p (or M_n) contain features associated with positive (or negative) sentiments.
- **Monotonicity constraints:**
  - $\sigma \leq \sigma' \Rightarrow \mu_{<\sigma,w>} \leq \mu_{<\sigma',w>}$, $w \in M_p$
  - $\sigma \geq \sigma' \Rightarrow \mu_{<\sigma,w>} \leq \mu_{<\sigma',w>}$, $w \in M_n$

Isotonic CRF

$$p(y|x) = \frac{1}{Z(x)} \exp\Big(\sum_{i} \sum_{\sigma,\tau} \lambda_{<\sigma,\tau>} f_{<\sigma,\tau>}(y_{i-1}, y_i) + \sum_{i} \sum_{\sigma,w} \mu_{<\sigma,w>} g_{<\sigma,w>}(y_{i-1}, x_i)\Big)$$

- "totally agree" is observed in the training data

- $\mu_{<PP,\ totally\ agree>} \geq \mu_{<NN,\ totally\ agree>}$

- We collect a lexicon compiled from MPQA (Wilson et al., 2005), General Inquirer (Stone et al., 1966), and SentiWordNet (Esuli and Sebastiani, 2006).

Training A Sentiment Classifier

- Authority and Alignment in Wikipedia Discussions (AAWD) corpus (Bender et al., 2011)

- 221 English Wikipedia discussions with positive and negative alignment annotations.

Roadmap

- A dispute corpus constructed from Wikipedia
- Online dispute detection
  - Sentence-level sentiment prediction
  - Dispute detection
- Conclusion

Online Dispute Detection

- **Sentiment features**
  - Sentiment distribution
    - $P(S)$, where $S \in \{NN, N, O, P, PP\}$
  - Sentiment transition distribution
    - $P(S_t \rightarrow S_{t+1})$, where $S_t, S_{t+1} \in \{NN, N, O, P, PP\}$
- **Two versions**
  - Global version: estimated from whole discussion
  - Local version: segment a discussion into three stages equally
    - For future work, we can leverage other topic segmentation techniques.

Online Dispute Detection

- Lexical Features
  - Unigram, bigram
- Topic Features
  - Category information
- Discussion Features
  - Number of turns
  - Number of participants
  - Average number of words in each turn

Experimental Setup

- Logistic regression
- Linear SVM
- RBF kernel SVM

- 5-fold cross-validation

Results

	Precision	Recall	F1	Accuracy
Baseline (Random)	50.0	50.0	50.0	50.0
Baseline (All dispute)	50.0	100.0	66.7	50.0
Logistic Regression	74.8	72.3	73.5	73.9
SVM + Linear	69.8	71.9	70.8	70.4
SVM + RBF	**77.4**	79.1	**78.3**	**80.0**

Results

	Precision	Recall	F1	Accuracy
Baseline (Random)	50.0	50.0	50.0	50.0
Baseline (All dispute)	50.0	100.0	66.7	50.0
Logistic Regression	74.8	72.3	73.5	73.9
SVM + Linear	69.8	71.9	70.8	70.4
SVM + RBF	**77.4**	79.1	**78.3**	**80.0**

Feature Analysis

	Precision	Recall	F1	Accuracy
Lexical	**75.86**	34.66	47.58	61.82
Topic	68.44	71.46	69.92	69.26
Discussion	69.73	**76.14**	**72.79**	71.54
Sentiment (Senti$_{g+l}$)	72.54	69.52	71.00	**71.60**

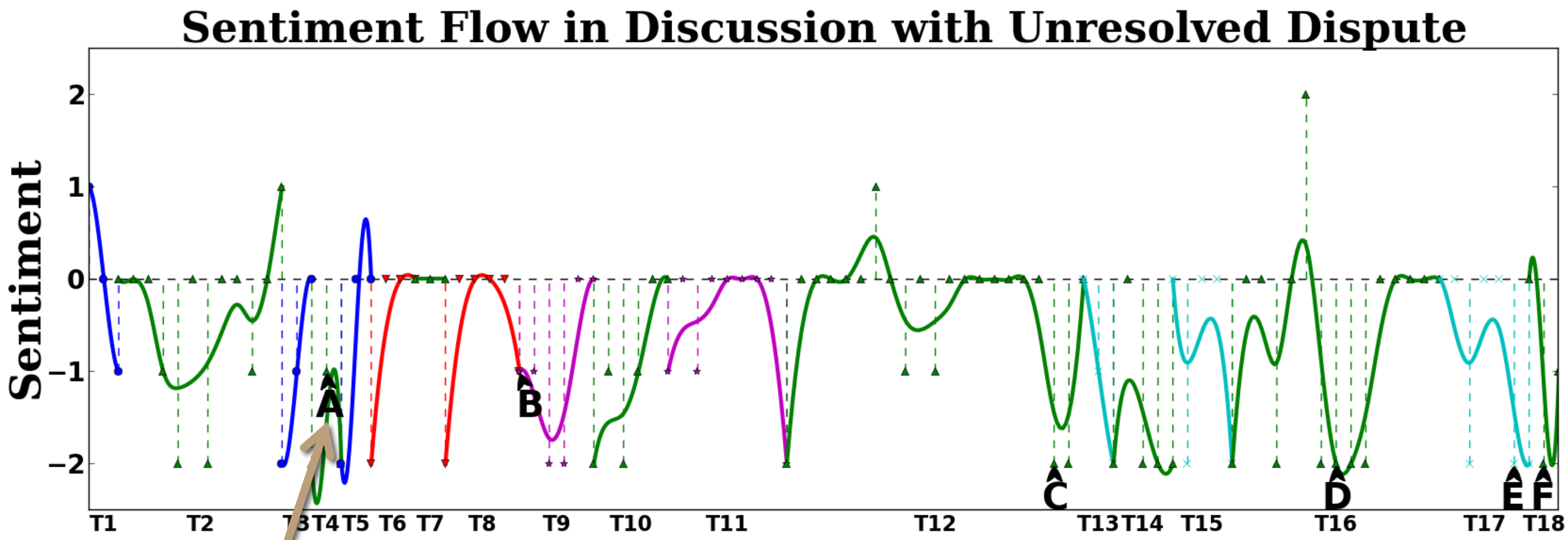
Feature Analysis

	Precision	Recall	F1	Accuracy
Lexical	75.86	34.66	47.58	61.82
Topic	68.44	71.46	69.92	69.26
Discussion	69.73	76.14	72.79	71.54
Sentiment (Senti$_{g+l}$)	72.54	69.52	71.00	71.60
Topic + Discussion	68.49	71.79	70.10	69.38
Topic + Discussion + Senti$_{g}$	77.39	78.36	77.87	77.74
Topic + Discussion + Senti$_{g+l}$	77.38	**79.14**	**78.25**	**80.00**
Lexical + Topic + Discussion + Senti$_{g+l}$	**78.38**	75.12	76.71	77.20

Feature Analysis

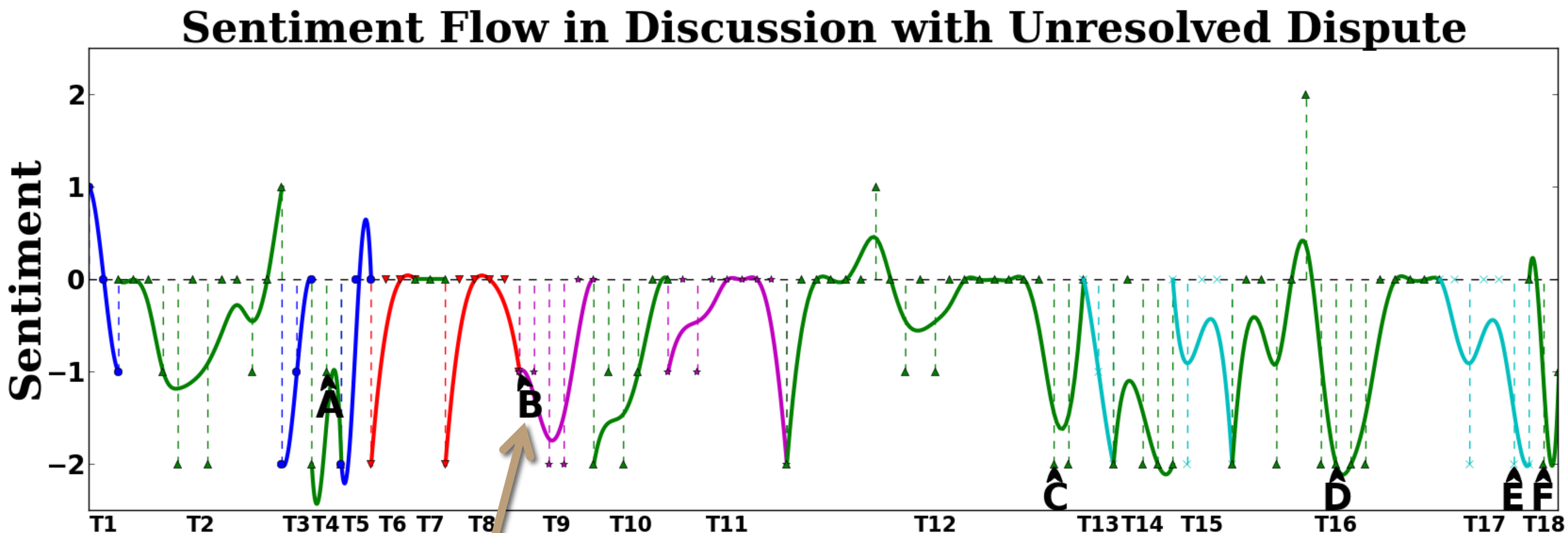
	Precision	Recall	F1	Accuracy
Lexical	75.86	34.66	47.58	61.82
Topic	68.44	71.46	69.92	69.26
Discussion	69.73	76.14	72.79	71.54
Sentiment (Senti$_{g+l}$)	72.54	69.52	71.00	71.60
Topic + Discussion	68.49	71.79	70.10	69.38
Topic + Discussion + Senti$_{g}$	77.39	78.36	77.87	77.74
Topic + Discussion + Senti$_{g+l}$	77.38	**79.14**	**78.25**	**80.00**
Lexical + Topic + Discussion + Senti$_{g+l}$	**78.38**	75.12	76.71	77.20

Visualization on Sentiment Flow



A: no, I sincerely plead with you... (N) If not, you are just wasting my time. (NN)

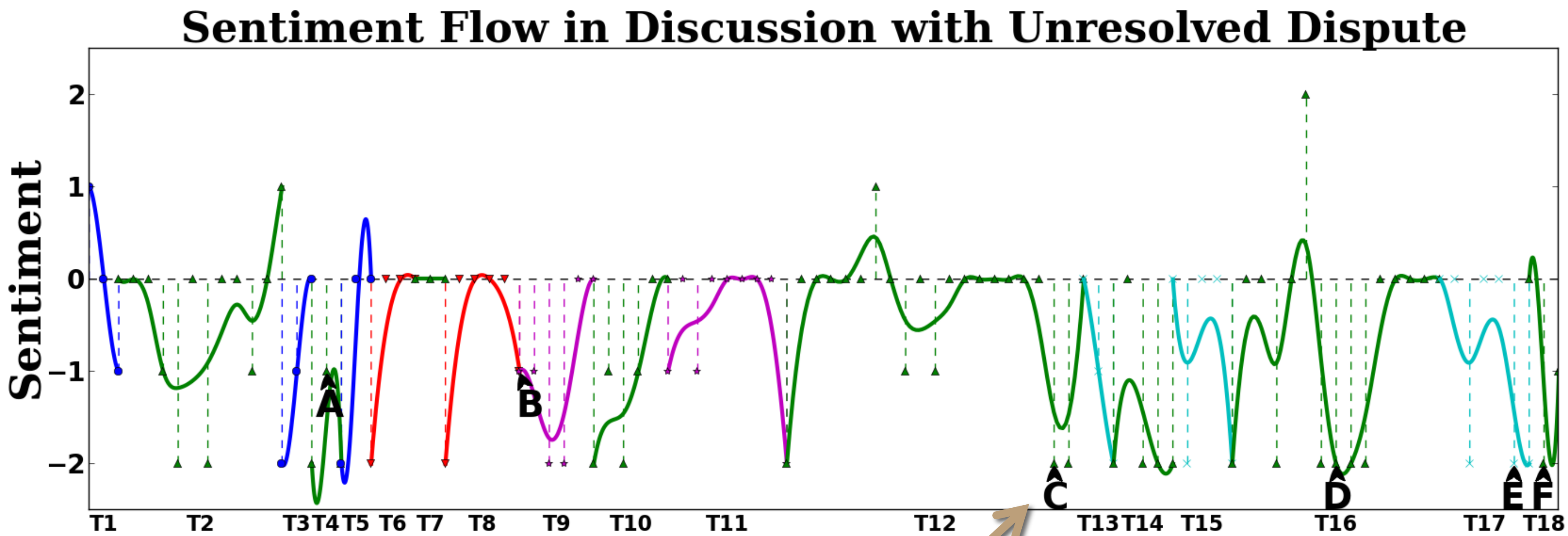
Visualization on Sentiment Flow



A: no, I sincerely plead with you... (N) If not, you are just wasting my time. (NN)

B: I believe Sweet's proposal... is quite silly. (NN)

Visualization on Sentiment Flow

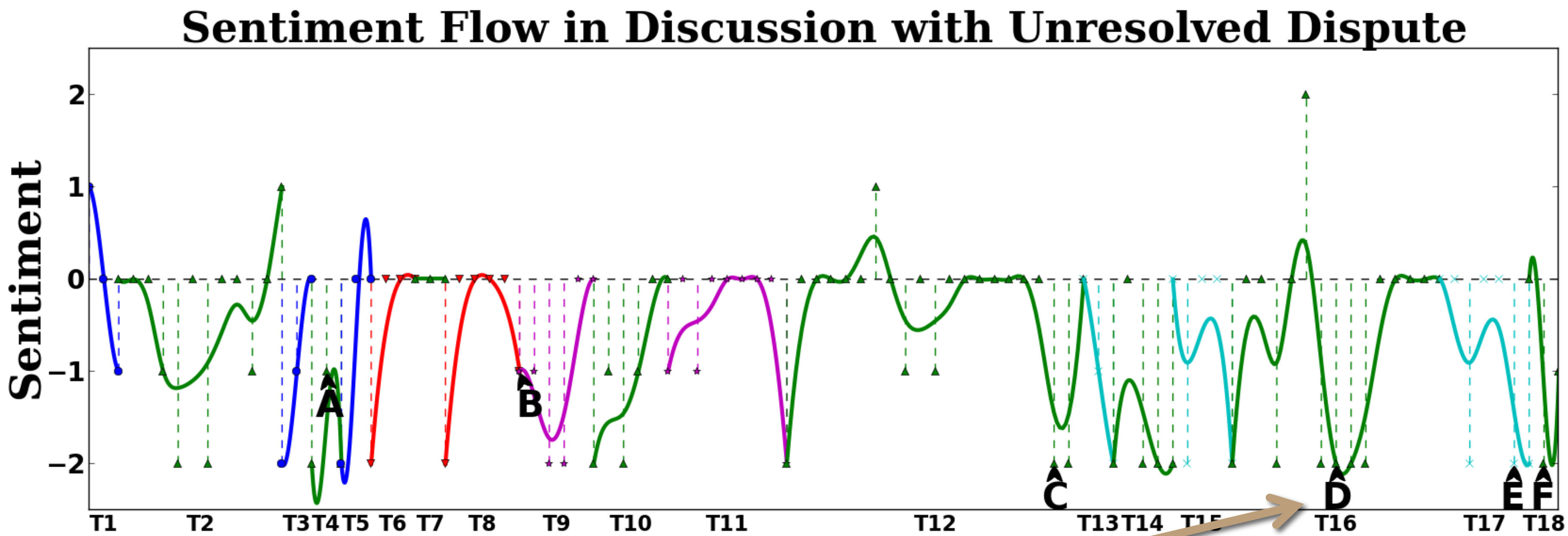


A: no, I sincerely plead with you... (N) If not, you are just wasting my time. (NN)

B: I believe Sweet's proposal... is quite silly. (NN)

C: Tell you what. (NN) If you can get two other editors to agree... I will shut up and sit down. (NN)

Visualization on Sentiment Flow



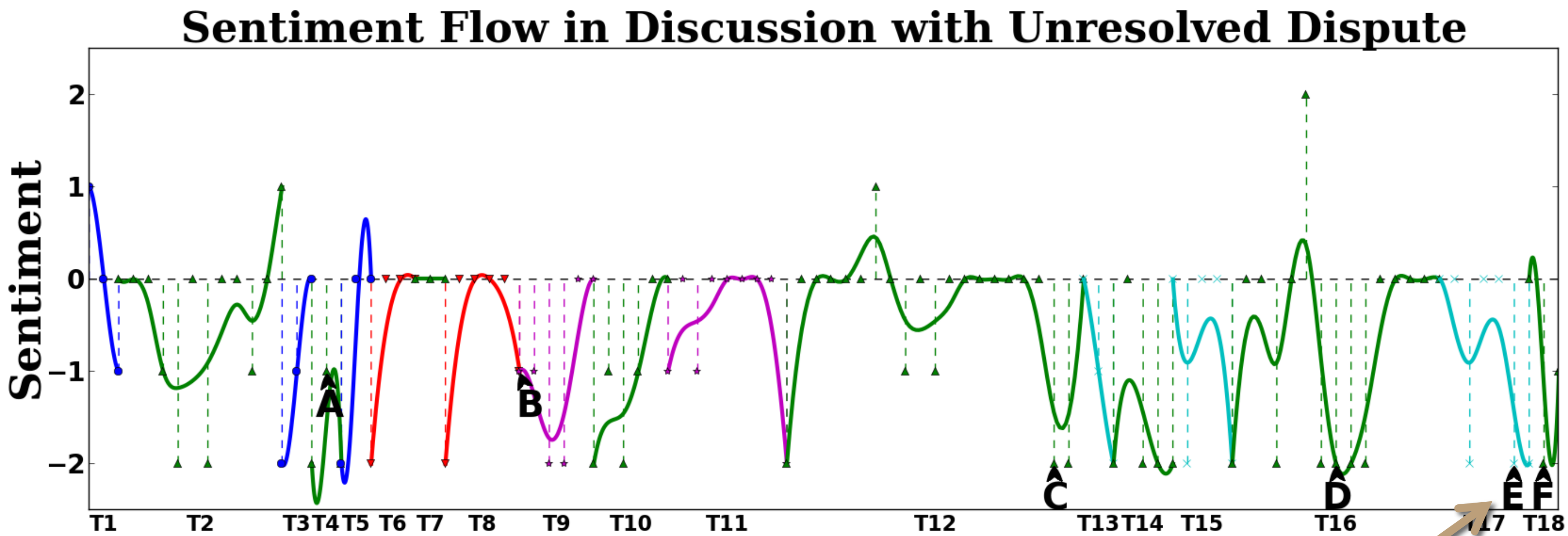
A: no, I sincerely plead with you... (N) If not, you are just wasting my time. (NN)

B: I believe Sweet's proposal... is quite silly. (NN)

C: Tell you what. (NN) If you can get two other editors to agree... I will shut up and sit down. (NN)

D: But some idiot forging your signature claimed that doing so would violate. (NN)... Please go have some morning coffee. (O)

Visualization on Sentiment Flow



A: no, I sincerely plead with you... (N) If not, you are just wasting my time. (NN)

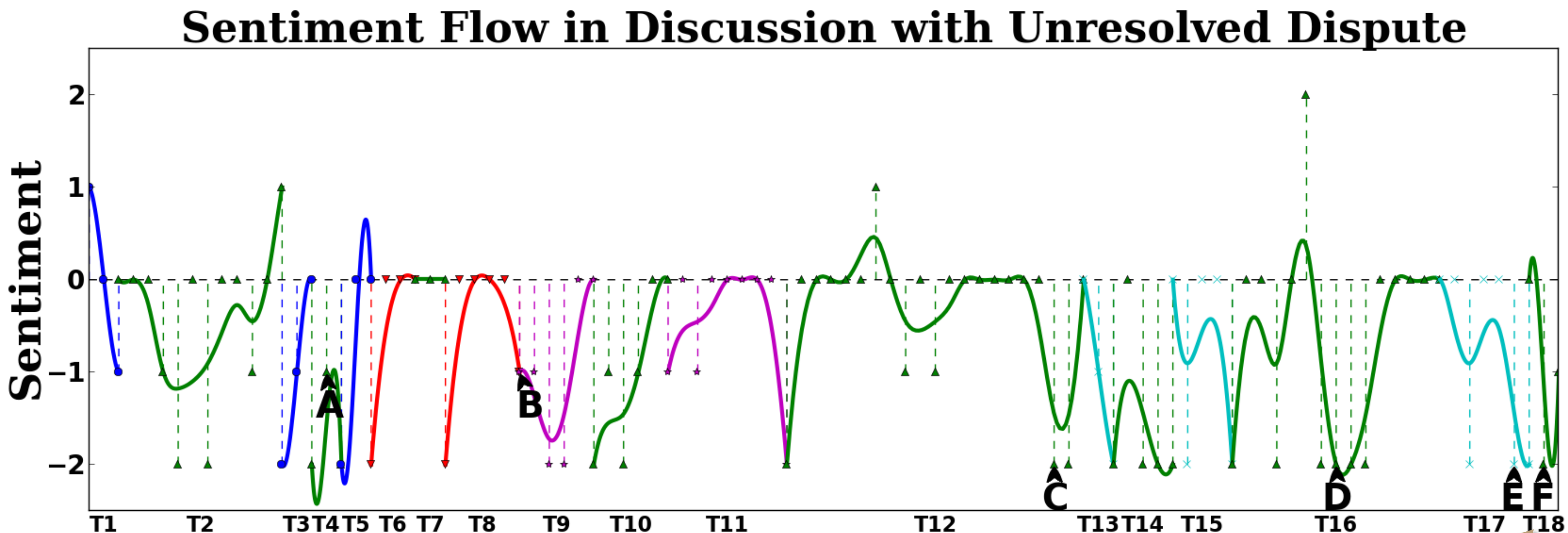
B: I believe Sweet's proposal... is quite silly. (NN)

C: Tell you what. (NN) If you can get two other editors to agree... I will shut up and sit down. (NN)

D: But some idiot forging your signature claimed that doing so would violate. (NN)... Please go have some morning coffee. (O)

E: And I don't like coffee. (NN) Good luck to you. (NN)

Visualization on Sentiment Flow



A: no, I sincerely plead with you... (N) If not, you are just wasting my time. (NN)

B: I believe Sweet's proposal... is quite silly. (NN)

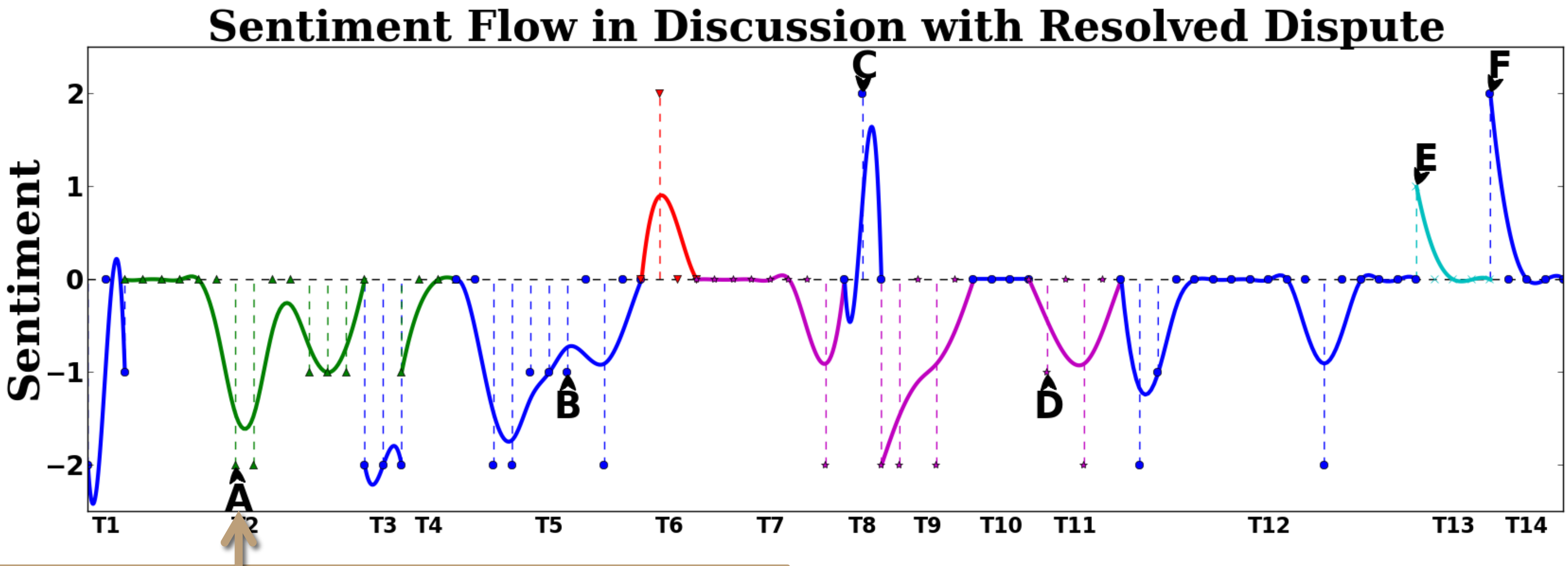
C: Tell you what. (NN) If you can get two other editors to agree... I will shut up and sit down. (NN)

D: But some idiot forging your signature claimed that doing so would violate... (NN)... Please go have some morning coffee. (O)

E: And I don't like coffee. (NN) Good luck to you. (NN)

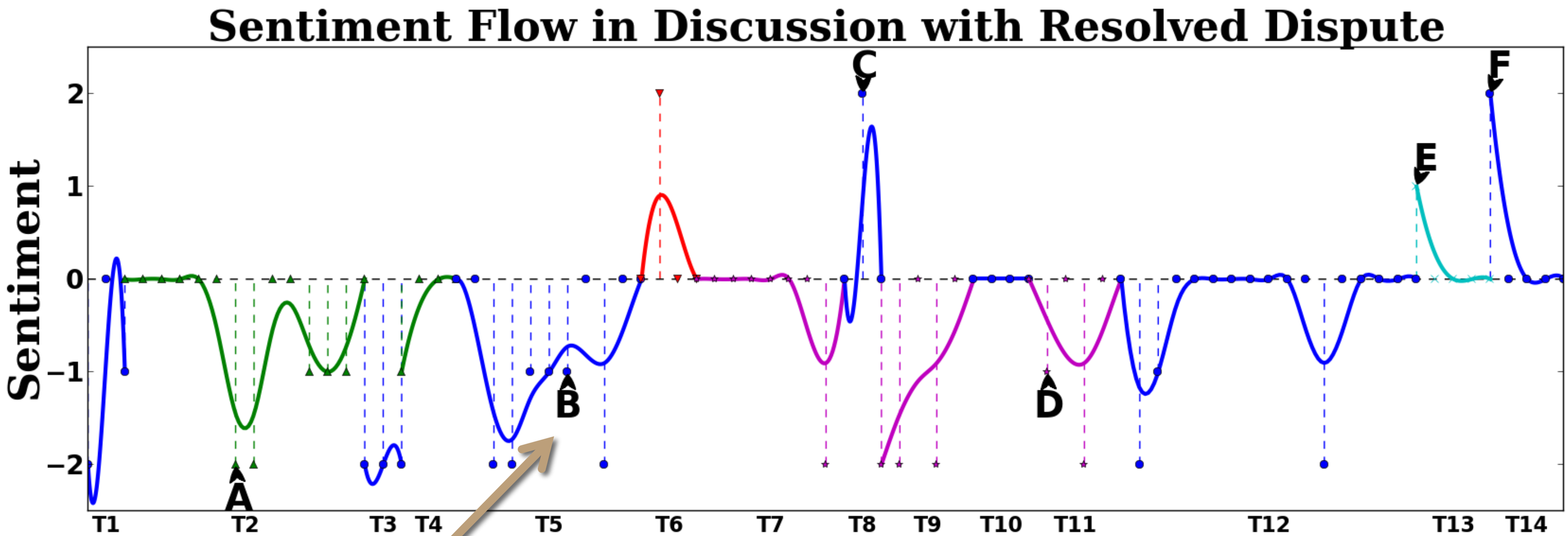
F: Was that all? (NN)... I think that you are in error... (N)

Visualization on Sentiment Flow



A: So far so confusing. (NN)...

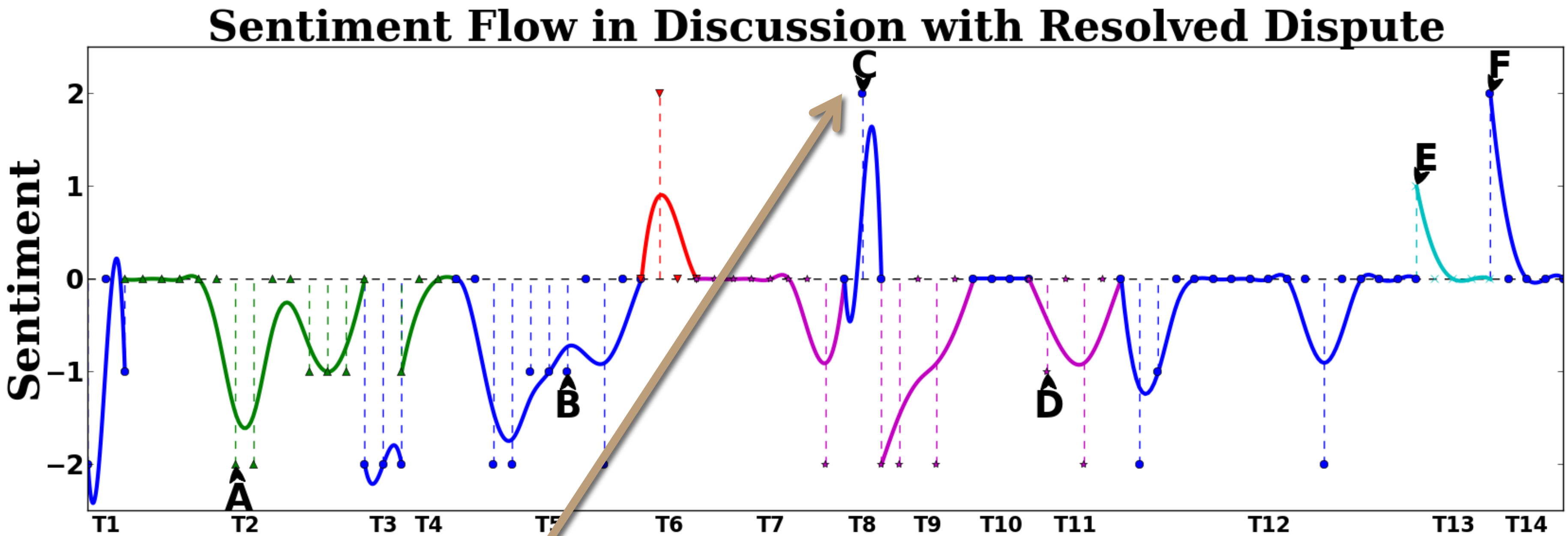
Visualization on Sentiment Flow



A: So far so confusing. (NN)...

B: ... I can not see a rationale for the landrace having its own article... (N) …

Visualization on Sentiment Flow

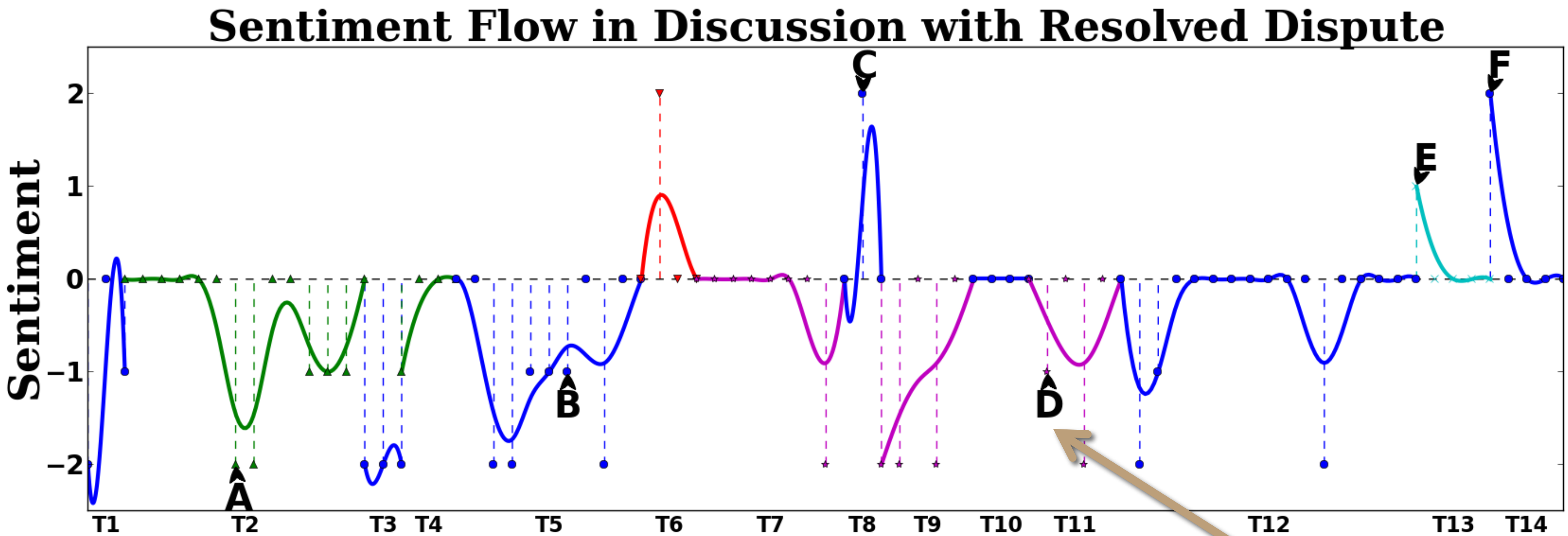


A: So far so confusing. (NN)...

B: ... I can not see a rationale for the landrace having its own article... (N) …

C: I've also copied your post immediately above to that article's talk page since it is a great "nutshell" summary. (PP)

Visualization on Sentiment Flow



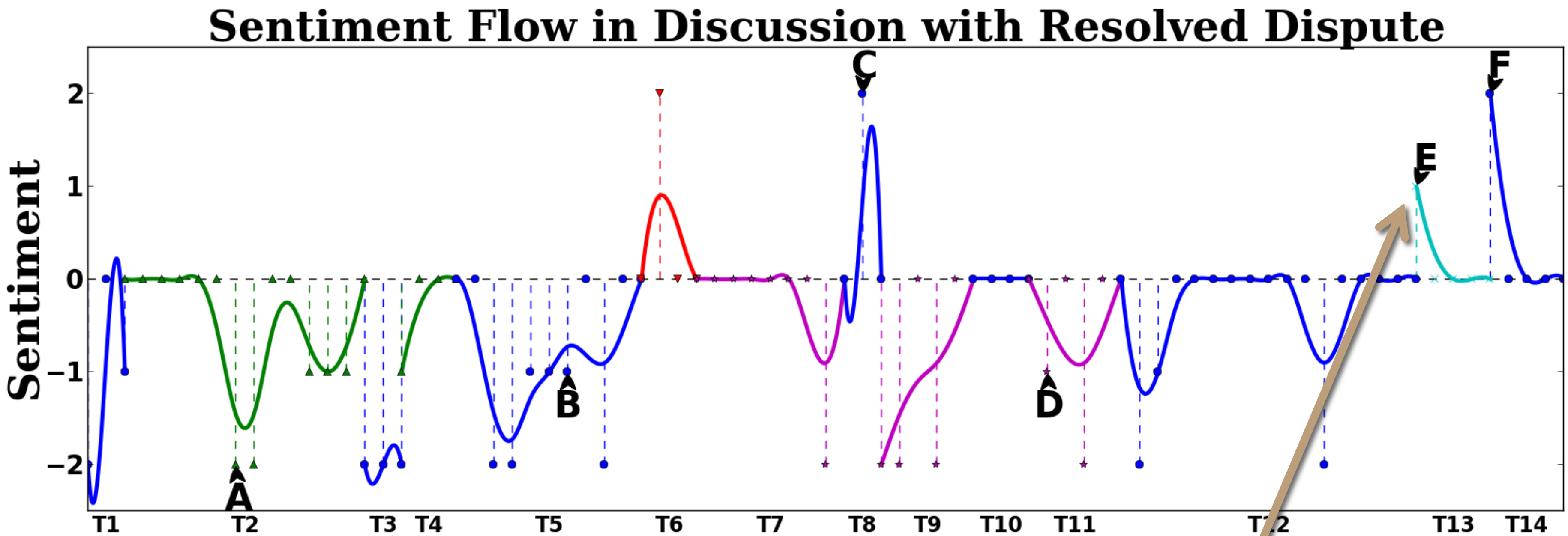
A: So far so confusing. (NN)...

B: ... I can not see a rationale for the landrace having its own article... (N) …

C: I've also copied your post immediately above to that article's talk page since it is a great "nutshell" summary. (PP)

D: Err.. how can the opposite be true... (N)

Visualization on Sentiment Flow



A: So far so confusing. (NN)...

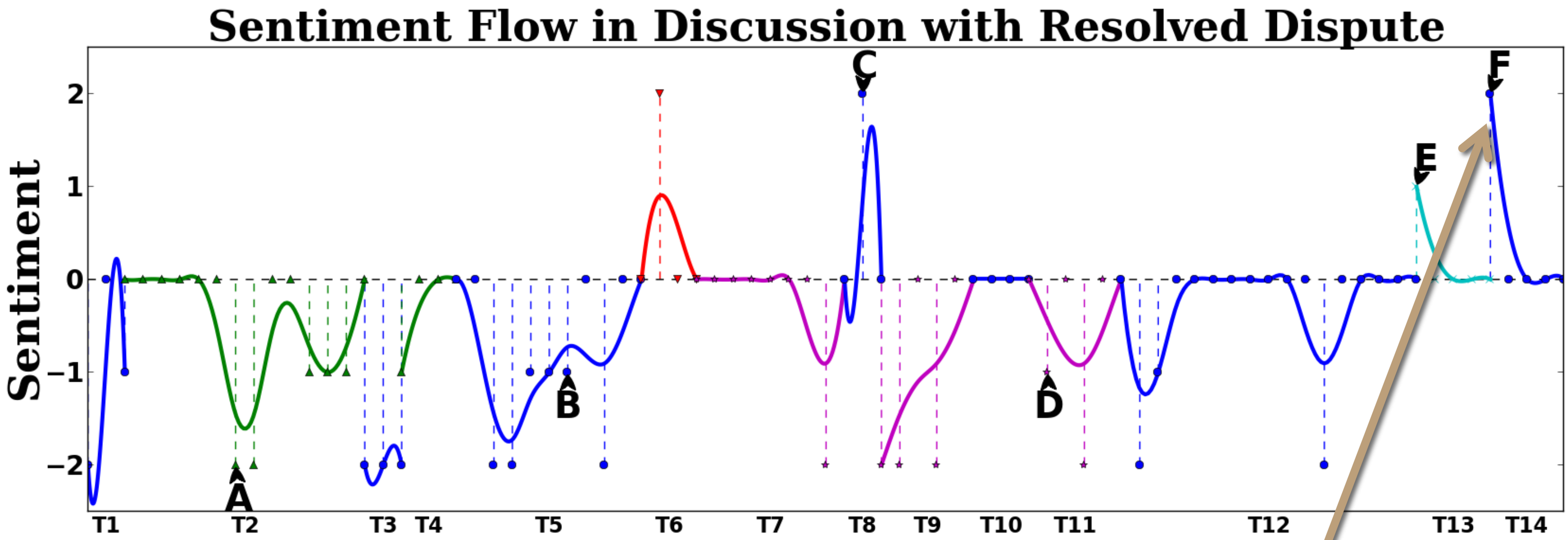
B: ... I can not see a rationale for the landrace having its own article... (N) …

C: I've also copied your post immediately above to that article's talk page since it is a great "nutshell" summary. (PP)

D: Err.. how can the opposite be true... (N)

E: Thanks for this, though I have to say some of the facts floating around this discussion are wrong. (P)

Visualization on Sentiment Flow



A: So far so confusing. (NN)...

B: ... I can not see a rationale for the landrace having its own article... (N) …

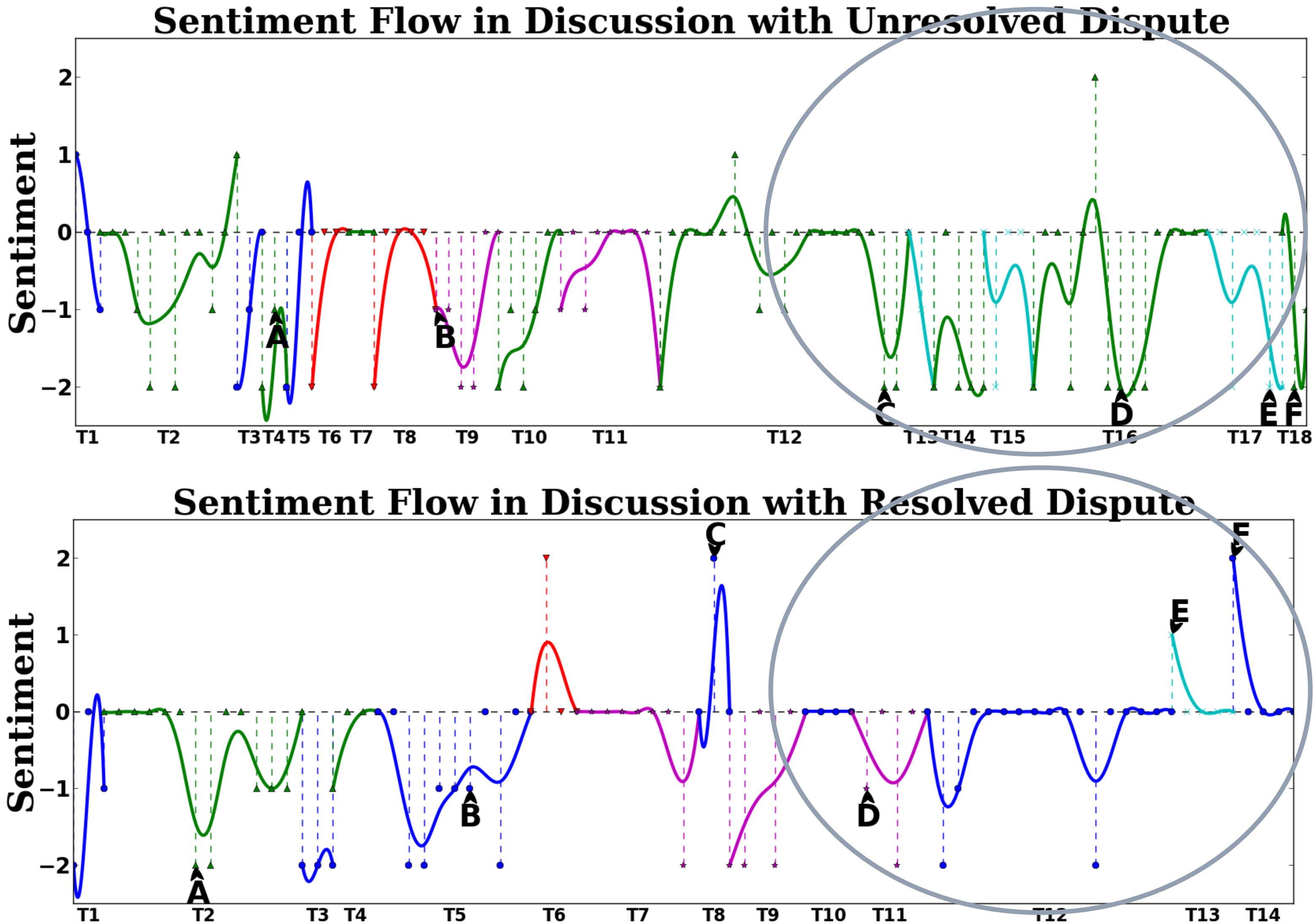
C: I’ve also copied your post immediately above to that article’s talk page since it is a great “nutshell” summary. (PP)

D: Err.. how can the opposite be true... (N)

E: Thanks for this, though I have to say some of the facts floating around this discussion are wrong. (P)

F: Great. (PP) Let’s make sure the article is clear on this. (O)

Visualization on Sentiment Flow



Discussions

- Dialog structure varies.
  - The recall for resolved dispute discussions is 0.86; and it is 0.78 for unresolved ones.
- The sentiment classifier has limitations.
  - *"I told you over and over again…"*: neutral or negative?
  - *"Wow, I am really enjoying this photography debate!"*: sarcasm is hard to detect.

Conclusion

- We present a sentiment analysis-based approach to online dispute detection.

- We create a dispute corpus from Wikipedia Talk pages to study the problem.

- Experiments demonstrate that classifiers trained with sentiment tagging features outperform others that do not.

Thank you!

Features for Sentence-Level Sentiment Prediction

- Lexical Features: unigrams/bigrams, number of words all uppercased, number of words
- Discourse Features: initial ngrams, repeated punctuations, number of negators
- Conversation Features: quote overlap with target, TFIDF similarity with target
- Sentiment Features: sentiment words

Evaluation on Sentiment Prediction

	Positive	Negative	Neutral
Baseline (Polarity)	22.53	38.61	66.45
Baseline (Distance)	33.75	55.79	88.97
SVM (3-way)	44.62	52.56	80.84
CRF (3-way)	56.28	56.37	89.41
CRF (5-way)	58.39	56.30	**90.10**
isotonic CRF	**68.18**	**62.53**	88.87