What to Do with an Airport? Mining Arguments in the German Online Participation Project Tempelhofer Feld

Matthias Liebeck, Katharina Esau, Stefan Conrad
Heinrich Heine University Düsseldorf

3rd Workshop on Argument Mining, ACL 2016, Berlin

Fachhochschule
für öffentliche Verwaltung
NRW

NRW-FORTSCHRITTSKOLLEG
ONLINE-PARTIZIPATION

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Outline

What is the Tempelhofer Feld?

- Former airport *Berlin-Tempelhof (THF)*, air traffic was ceased in 2008
- 300 hectare area that is mostly open space, used for recreation (inline skating, kite surfing, ...)

What is Online Participation?

- Involvement of citizens in relevant political or administrative decisions
- Cities offer their citizens an internet-based way to participate in drafting ideas for urban planning or in local political issues
- Examples:
  - Lörrach: sustainable urban development
  - Darmstadt and Bonn: gather proposals in participatory budgetings
  - Berlin: Tempelhofer Feld

Tempelhofer Feld + Online Participation

- Official online plattform[1] that includes citiziens in the planning of the area's future
- *ThF law* entered into force in 2014: structural changes are limited, for instance the construction of new buildings on the field is prohibited
- The project aims to collect ideas that improve the field for visitors while adhering to the ThF law.
- Official submission phase for proposals from November 2014 until the end of March 2015

[1] `https://tempelhofer-feld.berlin.de`

Tempelhofer Feld + Online Participation

- Forum-like plattform with proposals and comments
- Until July 2015, users proposed 340 ideas and wrote ≈1400 comments.
- ≈7000 sentences in the whole plattform
- Comments vary in length: on average 3.5 sentences

Desired System Output

- Assume that you have thousands of text comments
- An automatic extraction approach should answer three questions:
  1. What are suggestions that politicians can decide upon?
  2. What are reasons for/against the realization of these suggestions?
  3. How many citiziens express a pro/contra stance towards these suggestions?

Argumentation Model

- We tried to apply existing argumentation models, namely Toulmin and the claim-premise scheme.
- We quickly realized that
  - we have discourse between different users
  - attacks on logical conclusions are rather rare
  - users frequently express their wishes
  - users participate by providing reasons for and against other suggestions
  - suggestions cannot be classified as true or false
  - suggestions can be accepted without additional support

Argumentation Model

- We decided to modify the claim-premise family and its modification for persuasive essays from Stab and Gurevych 2014[2] to a three-part model in online participation processes: (i) major positions, (ii) claims, and (iii) premises
- **major positions**:
  - are options for actions or decisions that occur in the discussion (e.g., "*We should build a playground with a sandbox.*" or "*The opening hours of the museum must be at least two hours longer.*").
  - are most often someone's vision of something new or of a policy change.
  - In our practical view, major positions are unique suggestions from citizens that politicians can decide on.

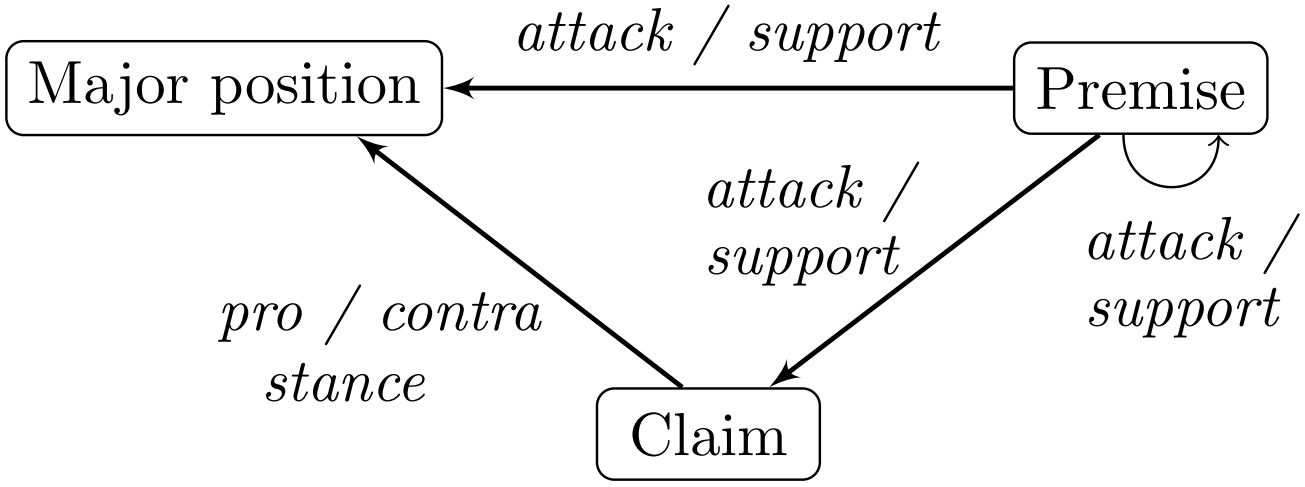
[2]Christian Stab and Iryna Gurevych. 2014. Annotating Argument Components and Relations in Persuasive Essays. COLING

Argumentation Model

- **claims**:
  - are pro or contra stances towards a major position (e.g. "*Yes, we should definitely do that!*")
- **premises**:
  - are defined as reasons that attack or support a major position, a claim or another premise
  - are used to make an argumentation comprehensible for others, by reasoning why a suggestion or a decision should be realized or why it should be avoided (e.g. "*This would allow us to save money.*")

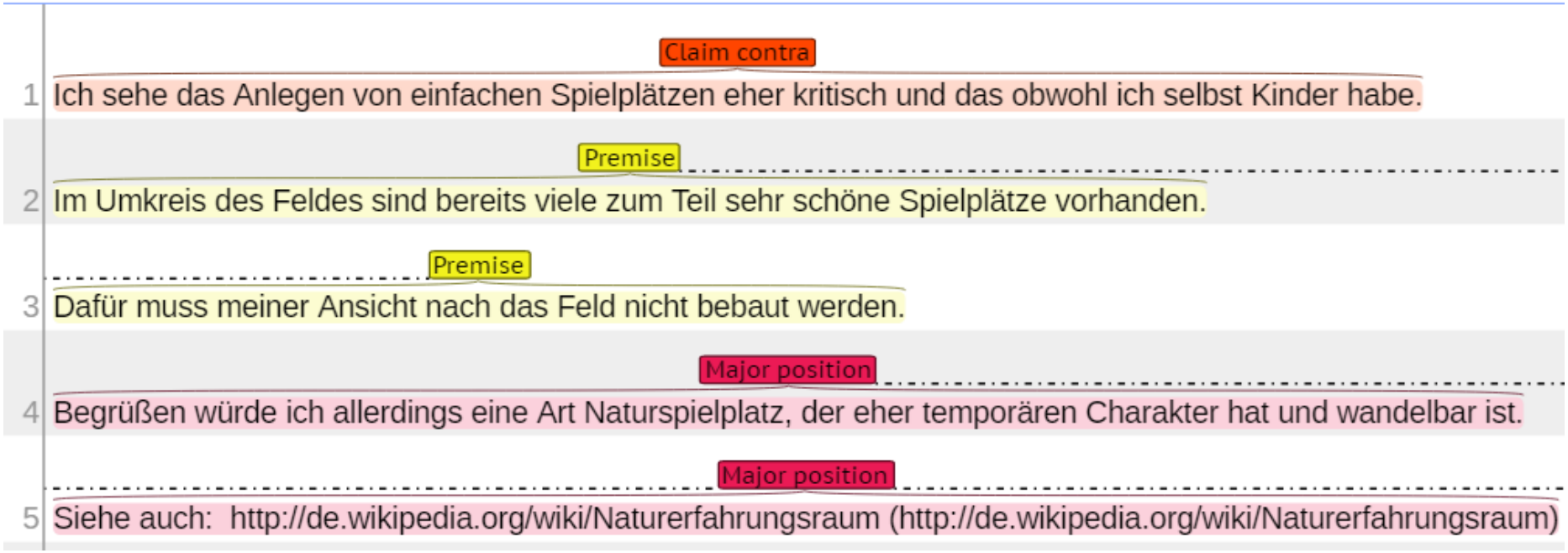
Argumentation Model

- Relations between the argument components



Annotation

- We used three annotators to annotate a subset of the online participation project.
- The annotators used *freely assignable spans*. Multiple annotations per sentence were possible.
- The dataset was annotated in the *brat rapid annotation tool*:



Inter-Annotator Agreement

- Before annotating our dataset, we took a subset (8 proposals and 74 comments, comprising 261 sentences and 4.1k tokens) to measure the IAA among the three annotators.
- We use *DKPro Agreement* to report our inter-annotator agreement values
  - Krippendorff's unitized alpha α_u
  - token-based observed agreement $A_{o,t}$
  - token-based Fleiss' kappa κ_t

  for the following three scenarios

  (i) joint measures over all categories
  (ii) category-specific values
  (iii) argumentative vs. non-argumentative units

Inter-Annotator Agreement

	$A_{o,t}$	κ_t	α_u
all	76.4	62.6	**78.0**
major positions	89.3	71.9	79.8
claims pro	96.3	66.1	59.0
claims contra	95.6	52.3	57.2
premises	80.9	61.5	80.1
AU / non-AU	90.7	49.1	**92.4**

- Reliable agreement between our three annotators:
  - $\alpha_u = 0.924$ for argumentative versus non-argumentative spans
  - $\alpha_u = 0.78$ for the joint measure for all categories

Statistics of the annotated corpus

- 72 proposals
- 575 comments
- $\approx$2400 sentences
- 88% of the tokens belong to argumentative spans
- 3.6% of the sentences were annotated with more than one argument component

Classification

- Preprocessing:
  - *OpenNLP*: Sentence splitting and tokenization
  - *Mate Tools*: POS-tagging and dependeny parsing
- Classification tasks:
  - **Subtask A**: Classify sentences as argumentative or non-argumentative
  - **Subtask B**: Classify argument components in argumentative sentences with exactly one annotated argument component
- Training/Test data: 80% training set, 20 % test data
  - **Subtask A**: ≈2000 sentences for training
  - **Subtask B**: ≈1600 sentences for training

Features

All features are sentence-based:

- n-Grams
  - Unigrams
  - Bigrams
- Grammatical features:
  - L_2-normalized POS-Tag distribution
  - L_2-normalized dependency distribution
- Structural features:
  - token count
  - comma count / token count
  - dot count / token count
  - last-token of a sentence as a one-hot encoding (‘.’, ‘!’, ‘?’, ‘OTHER’)
  - number of links

Classification

- We evaluated three classifiers:
  - SVM with an RBF kernel
  - Random forest
  - k-nearest neighbor
- Gridsearch with 10-fold cross validation on the training set
- Evaluation metric: macro-averaged F_1
- We evaluated different feature combinations and report their results:

Results

Feature Set	AU / non-AU			Argument Components		
	SVM	RF	k-NN	SVM	RF	k-NN
Unigram	65.99	68.13	61.00	64.40	59.41	40.30
Unigram,lowercased	66.69	64.53	62.26	65.32	53.35	38.25
Bigram	41.79	50.48	16.25	46.62	50.42	11.51
Grammatical	55.88	52.24	48.52	59.54	47.89	46.81
U + G	**69.77**	58.39	64.87	**68.50**	57.13	35.90
U + G + Structural	67.50	61.14	54.07	65.99	59.46	47.27

Table: Macro-averaged F_1 scores for the two classification problems: (i) classifying sentences as argumentative and non-argumentative, (ii) classifying sentences as major positions, claims, and premises.

Results

- The best results for both subtasks were achieved by an SVM and *unigrams + grammatical features*.
- Confusion matrix for the best approach of subtask B:

Actual \ Predicted	MP	C	P	$\sum$
MP	63	4	43	110
C	9	48	12	69
P	27	20	172	219
$\sum$	99	72	227	398

- The classification of premises works well.
- Major positions are often misclassified as premises.

Conclusion

- New corpus for German argumentation mining (will be released shortly)
- Argumentation model for online participation
- Inter-annotator agreement study
- Two classification tasks:
  - We evaluated different feature combinations and multiple classifiers.
  - The best results of 69.77% in subtask A and 68.5% in subtask B were both achieved by a support vector machine.

Future Work

- Additional features to further increase our classification results
- Automatically detect tokens that form a group, based on the content. For this, we could use the token-based BIO scheme, which divides tokens into beginning (B), inner (I), and other (O) tokens of an argument component
- Identify more freely available corpora for online participation to which we can apply our model for a comparative study

Thanks for listening!

References

- Google Maps: Bilder (c) 2018 Google, Kartendaten (c) 2018 GeoBasis-De/BKG ((c)2009), Google