Computational Argumentation — Part V

Argument Mining

Henning Wachsmuth

henningw@upb.de

PADERBORN UNIVERSITY

Learning goals

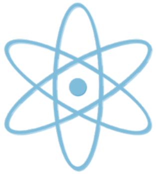
- **Concepts**
  - Major argument models
  - Definitions, goals, and tasks in argument mining

https://commons.wikimedia.org

- **Methods**
  - Filtering of argumentative texts
  - Segmentation of argumentative and non-argumentative units
  - Classification of types of argumentative units
  - Identification of relations between units and arguments

https://pixabay.com

- **Associated research fields**
  - Argumentation theory
  - Natural language processing



https://pixabay.com

- **Within this course**
  - The first of three main stages in computational argumentation

Outline

Argumentative discourse units (recap)

- **Argumentative function**
  - Argumentative language supports or attacks stances on controversial issues.
  - Any claim, or reason for a claim, has an argumentative function.
- **Argumentative unit** (aka argument component)
  - A contiguous text span with a specific argumentative function, demarcated by neighboring spans with a different function
- **Argumentative discourse unit (ADU)**
  - An argumentative unit, or a non-argumentative text span, that has a rhetorical or dialectical function, gives background information, ...

    Some literature sees only argumentative units as ADUs.

non-argumentative — argumentative

” If you wanna hear my view, I think that the EU should allow rescue boats in the Mediterranean Sea. Many innocent refugees will die if there are no rescue boats. Nothing justifies to endanger the life of innocent people.”

argumentative — argumentative

Arguments (recap)

- **Argument**
  - A composition of a set of argumentative units, where one takes the role of a *conclusion* and each other the role of a *premise*
  - Conclusion. A claim that conveys a stance on a controversial issue, implicitly or explicitly
  - Premise. A reason given to support (or object to) the truth of the claim

Conclusion / Premises

Conclusion	*The EU should allow rescue boats in the Mediterranean Sea.*
Premise 1	*Many innocent refugees will die if there are no rescue boats.*
Premise 2	*Nothing justifies to endanger the life of innocent people.*

- **Arguments are inherently relational**
  - An argument defines a relation where premises support a conclusion.
  - A premise may also serve as a counterconsideration that objects to a conclusion. It is then usually *undercut* in the same argument.

What is argument mining?

- **Argument mining** (aka argumentation mining)
  - The identification of argumentative structure in natural language text, in terms of units and their relations

    Can be seen as the core NLP task in computational argumentation
  - May be based on different argument models
  - Often, the argument mining process includes multiple steps.

non-argumentative

argumentative

Conclusion

" If you wanna hear my view, I think that the EU should allow rescue boats in the Mediterranean Sea. Many innocent refugees will die if there are no rescue boats. Nothing justifies to endanger the life of innocent people."

support

Premise

support

Premise

- **Why argument mining?**
  - Real-world arguments are often "hidden" in longer text, possibly fragmented.
  - Mining provides the basis for any argument analysis and any application.

    Exception: Arguments, and their structure, are already given in the source data.

Overview of the argument mining process

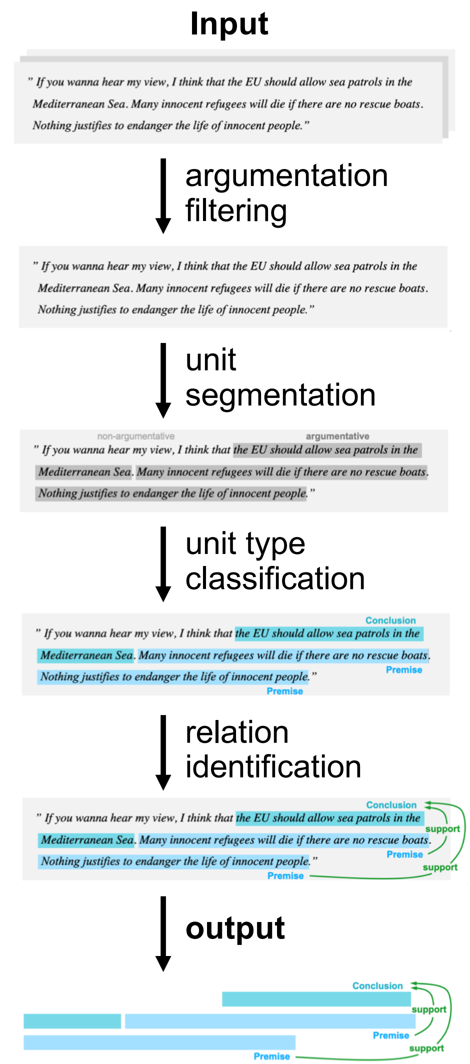
- **General process signature**
  - Input. A set of (plain) texts
  - Output. The argumentative structure of each text

    What structure is mined exactly depends on the employed argument model.

- **Main high-level tasks**
  - Argumentation filtering. Finding argumentative texts
  - Unit segmentation. Finding argumentative units
  - Unit type classification. Finding types of units
  - Relation identification. Finding relations between units

- **Notice**
  - Different task decompositions exist in literature.

    Some also see assessment tasks as part of argument mining.
  - The tasks are not always tackled in the given ordering.
  - Not all tasks always need to be tackled.



Next section: Argument models

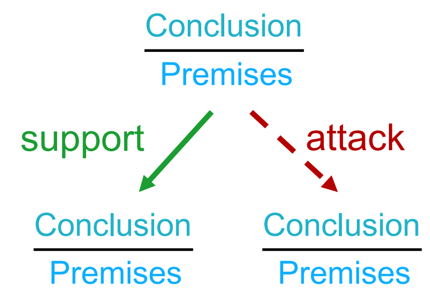
What is an argument model?

- **Argument model**
  - Formalized definition of the concepts distinguished for an argument
  - Concepts usually reflect structural and/or semantic aspects.
  - Used in computational argumentation to operationalize argument processing

    The concepts define the types of meta-information created by mining (and partly assessment) methods.

- **Argument model as a graph**
  - Most models can be represented as a graph *G = (V, E)* with nodes *V* and edges *E.*
  - Labeling/Weighting functions may be given for *V* and *E.*

    We will mostly just use visual representations of the graph elements.



- **Argument models**
  - Several models of arguments have been proposed in argumentation theory and computational research from practice.
  - What model to use depends on the given genre and intended application, i.e., on what the distinguished concepts are meant to be used for.

Overview of existing argument models

- **Models from theory**
  - Toulmin model. Fine-grained unit roles (Toulmin, 1958)
  - Freeman model. Dialectical exchange of views (Freeman, 2011)
  - Argumentation schemes. Form of inference within an argument (Walton et al., 2008)
  - IBIS. Relations between issues, stances, and arguments (Kunz and Rittel, 1970)
  - Weighted bipolar argumentation. Support and attack of weighted arguments (Amgoud and Ben-Naim, 2018)
  - Abstract argumentation framework. Attacks between arguments (Dung, 2015)

  ... among others

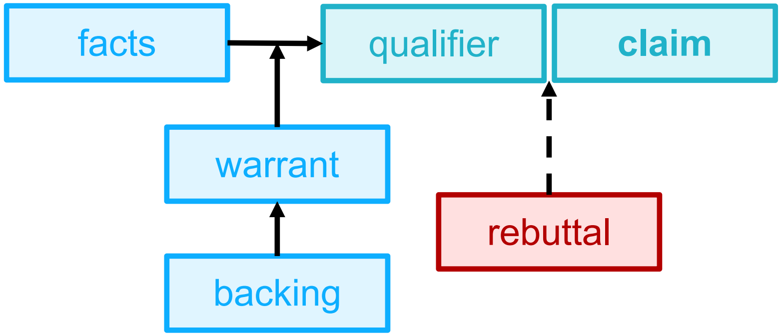
- **Models from practice**
  - Essay-specific. Hierarchical relations of claims and premises (Stab, 2017)
  - Editorial-specific. Fine-grained strategy-related unit roles (Al-Khatib et al., 2016)
  - Robust models. e.g., claim on issue, with stance and evidence (Bar-Haim et al., 2017)

  ... among others

- **Notice**
  - Some theoretical models are also used in practice, partly in simplified form.

Toulmin's argument model

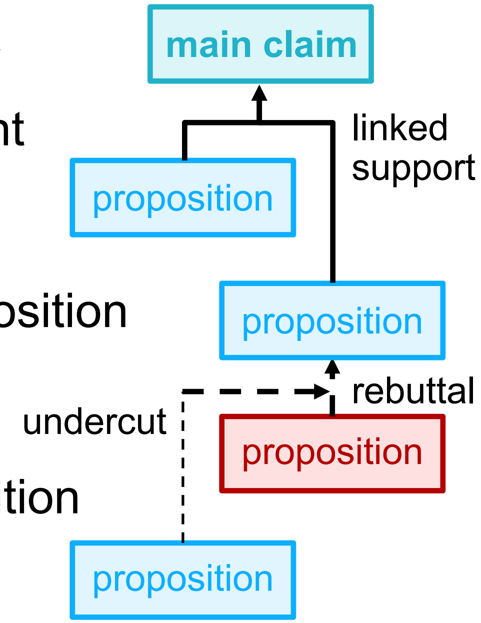
- **Toulmin model** (Toulmin, 1958)
  - Captures an argument's internal structure with fine-grained unit roles
    The relation between the roles is clear by definition.



- **Unit types**
  - Claim. A conclusion, as defined above
  - Qualifier. Constraint or uncertainty of the claim
  - Facts (aka data/grounds). Evidence given to support the claim
  - Warrant. Defeasible rule for why the claim can be inferred from the facts
  - Backing. Justification of the warrant
  - Rebuttal. Circumstances under which the claim (or warrant) does not hold
    Backing, qualifier, and rebuttal are optional.
- **Discussion**
  - The model clarifies how arguments work, but few real-life arguments match it.
    Units such as the warrant are often left implicit. Also, units may mix up more than one role.
  - Simplified variants have been used in practice. (Habernal and Gurevych, 2016)

Freeman's argument model based on Peldszus and Stede (2013)

- **Freeman model** (Freeman, 2011)
  - Captures the (hypothetical) dialectic exchange in an argument between a proponent defending a claim and an opponent attacking it
- **Unit types**
  - (Main) Claim. The proposition the proponent argues for
  - Proposition. Any other unit of the proponent or opponent
- **Relation types**
  - (Linked) Support. Inference from proposition(s) to proposition
    Peldszus and Stede (2013) consider *example* as a special type of support.
  - Rebuttal. Attack of the acceptability of a proposition
  - Undercutter. Attack of the inference based on a proposition

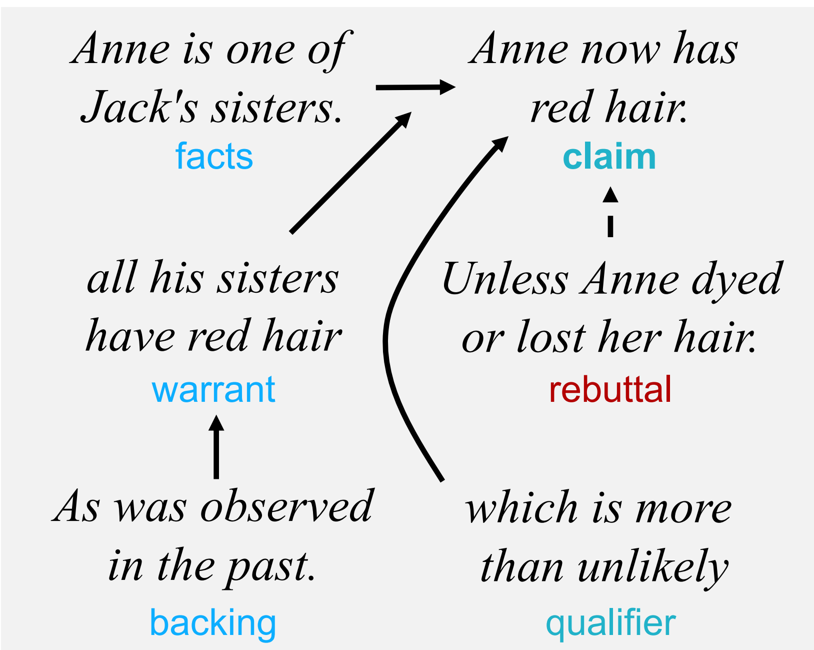


- **Discussion**
  - Freeman aimed to integrate Toulmin's ideas with informal logic.
  - In practice, a robust model at least for "clean" arguments (Peldszus and Stede, 2015)

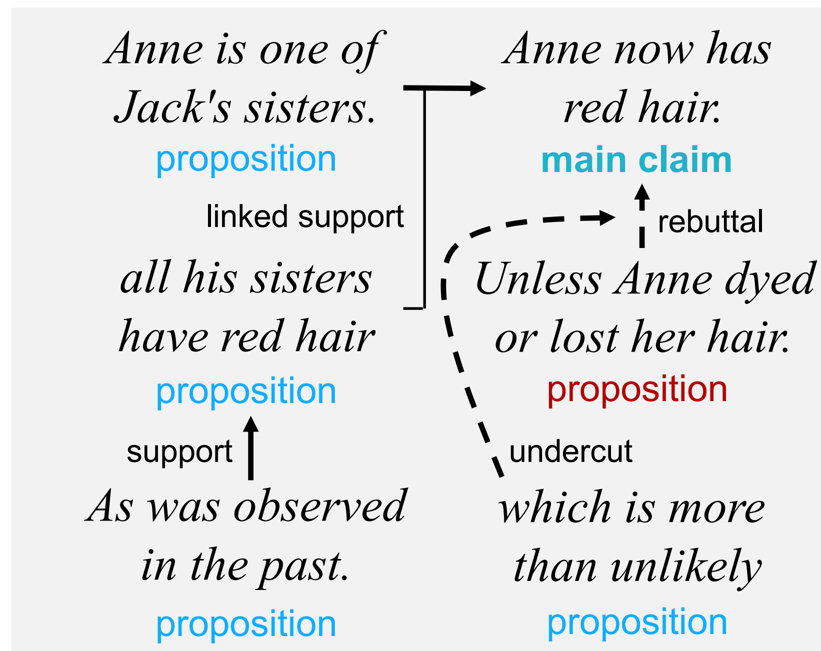
Examples: Toulmin vs. Freeman model

Anne is one of Jack's sisters. As was observed in the past, all his sisters have red hair. Unless Anne dyed or lost her hair, which is more than unlikely, Anne now has red hair.

- **Toulmin model**



- **Freeman model**

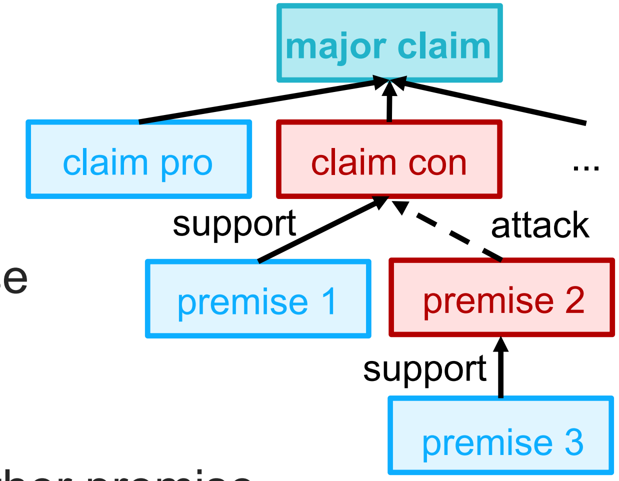


Essay-specific argument model

- **Essay-specific model** (Stab, 2017)
  - Captures the hierarchical structure of monological argumentative text

- **Unit types**
  - Major claim. The thesis of the text
  - Claim. The conclusion of an argument; has a stance towards the thesis
  - Premise. The premise of a claim or other premise

  Maximum one claim per paragraph



- **Relation types**
  - Support. The support of a claim/premise by another premise
  - Attack. Analog for attacks

  Relations do not cross paragraph boundaries.

- **Discussion**
  - Tuned towards the characteristics and conventions of persuasive essays
  - The assumptions behind may not generalize to many genres

Editorial-specific argument model

- **Editorial-specific model** (Al-Khatib et al., 2016)
  - Captures fine-grained unit roles related to an author‘s argumentation strategy
- **Unit types**
  - Assumption. Claim, fact, or similar that requires justification
  - Common ground. Self-evident fact, accepted truth, ...
  - Anecdote. Example, personal experience, specific event, ...
  - Testimony. Reference to a statement of an expert, authority, ...
  - Statistics. Reference to a finding from a study or similar
  - Other. Any other unit

- **Discussion**
  - Tuned towards the characteristics of news editorial argumentation
  - The types encode meaning rather than structural information.
  - The evidence types included are adopted from other models. (Rinott et al., 2015)
  - The ”assumption“ type could benefit from further decomposition.

Examples: Essay and editorial argumentation

- **Essay argumentation** (Stab, 2017)

claim pro *Living and studying overseas is an irreplaceable experience when it comes to learn standing on your own feet*

attack

support

premise *One who is living overseas will of course struggle with loneliness, living away from family and friends*

the one will learn living without depending on anyone else premise

attack

premise *those difficulties will turn into valuable experiences in the following steps of life*

- **Editorial argumentation** (Al-Khatib et al., 2016)

assumption

[...] It seems a college education is part of the American dream that's easy to buy (or borrow) into, but hard to pay off.

assumption

With tuition soaring, and the middle class shrinking along with their incomes, many students and their families are left holding incredibly expensive bags. In 2013, 69% of graduating seniors at public and private nonprofit colleges took out student loans to pay for college, and ”about one-fifth of new graduates' debt was in private loans,“ according to the Project on Student Debt. [...]

statistics

testimony

Yet the party line that college education is the middle class' only hope for upward mobility persists - it will even be the message of President Obama's last stop on his ”SOTU Spoiler“ tour in Knoxville, Tennessee. [...] assumption anecdote

Next section: Argumentation filtering

I. Introduction to computational argumentation

II. Basics of natural language processing

III. Basics of argumentation

IV. Argument acquisition

V. Argument mining

- a) Introduction
- b) Argument models
- **c) Argumentation filtering**
- d) Unit segmentation
- e) Unit type classification
- f) Relation identification
- g) Conclusion

VI. Argument assessment

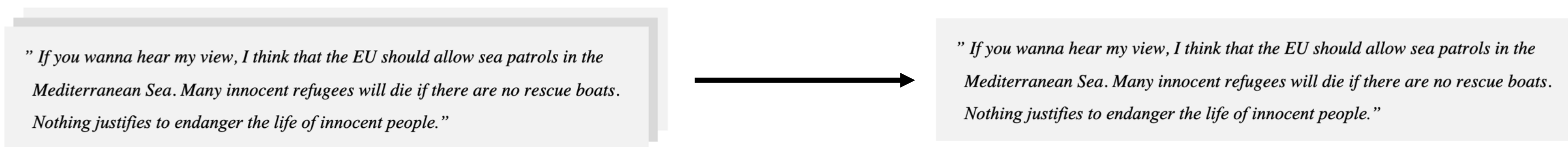
VII. Argument generation

VIII. Applications of computational argumentation

IX. Conclusion

What is argumentation filtering?

- **Argumentation filtering**
  - The retrieval of all argumentative texts from a given text collection
  - Input. Any set of texts
  - Output. The subset of texts considered to be argumentative



- **Exemplary approaches**
  - Ranking Wikipedia articles using argumentativeness lexicon (Roitman et al., 2016)
  - Classifying persuasive web texts using word *n*-grams (Habernal and Gurevych, 2017)

- **Discussion**
  - Few approaches specifically tuned towards argumentation exist.
  - The filtering can be seen as a specific genre classification task, for which extensive research exist. (out of scope here)

Filtering argumentative Wikipedia articles (Roitman et al., 2016)

- **Task**
  - Given the whole Wikipedia and a controversial issue, return the top *k* articles with the most argumentative potential for the issue.
- **Approach**
  - Rank all articles by relevance to the issue.
  - Score argumentativeness using features derived from a controversy lexicon.
  - Rerank articles by argumentativeness.

dispute, disputable, disagreement, debate, polemic, feud, question, schism, wrangle, controversy, dispeace, dissension, criticism, argue, disagree, argument, claim, conflict, opposition, adversary, antagonism, oppose, object, loggerheads, quarrel, fuss, moot, hassle, altercate, case, evidence, clash, issue, problem, emphasize, recommend, suggest, assert, defend, maintain, reject, support, challenge, doubt, refute, confirm, prove, validate, establish, substantiate, verify, against, resist, support, agree, consent, concur, accept, refuse, plead, right, justify, justification

- **Data**
  - 1739 claims in 626 out of 3M articles
- **Results**
  - Recall of claims

Approach	R@5	R@10	R@20
Issue relevance only	0.13	0.25	0.34
Issue relevance + argumentativeness	0.28	0.41	0.51

Next section: Unit segmentation

I. Introduction to computational argumentation

II. Basics of natural language processing

III. Basics of argumentation

IV. Argument acquisition

V. Argument mining

VI. Argument assessment

VII. Argument generation

VIII. Applications of computational argumentation

IX. Conclusion

a) Introduction
b) Argument models
c) Argumentation filtering
d) Unit segmentation
e) Unit type classification
f) Relation identification
g) Conclusion

What is unit segmentation?

- **Unit segmentation**
  - The segmentation of a text into ADUs, i.e., argumentative units and their non-argumentative counterparts
  - Input. Usually, a plain text (often assumed to be argumentative)
  - Output. All ADUs in the text, defined by their character/token boundaries

non-argumentative **argumentative**

" If you wanna hear my view, I think that the EU should allow rescue boats in the Mediterranean Sea. Many innocent refugees will die if there are no rescue boats. Nothing justifies to endanger the life of innocent people."

- **How to model that computationally?**
  - Individual classification of candidate start/end character/token boundaries
  - Sequence labeling in terms of BIO tagging of each token in a text (see below)

  ... along with some variations

Example: Unit segmentation of an essay paragraph

- **How good are humans in unit segmentation?**
  - Given the following essay paragraph on ”living overseas“, identify the ADUs.

> ” *Living and studying overseas is an irreplaceable experience when it comes to learn standing on your own feet. One who is living overseas will of course struggle with loneliness, living away from family and friends but those difficulties will turn into valuable experiences in the following steps of life. Moreover, the one will learn living without depending on anyone else.*

example from Stab and Gurevych (2014a)

- **What makes unit segmentation challenging?**
  - What is argumentative may depend on the issue being discussed.
  - Even humans may disagree on the correct segmentation.
  - No clear general definition exists of what makes up the boundaries of ADUs.
    Often, an ADU is a clause or sentence w/o discourse markers, but multiple-sentence ADUs exist (Rinott et al., 2015).

Overview of units and their segmentation

- **Argumentative units across genres**
  - Some genres are very dense in terms of argumentative units.
  - Others have a low proportion only, or argumentativeness is issue-dependent.

- **Argumentative units in selected genres**
  - Persuasive essays. Nearly everything is argumentative. (Stab and Gurevych, 2014a)
  - News editorials. Many ADUs rather have a *rhetorical* role. (Al-Khatib et al., 2016)
  - Wikipedia articles. Argumentativeness is issue-dependent. (Rinott et al., 2015)
  - Forum discussions. Argumentativeness strongly varies. (Habernal and Gurevych, 2017)

- **Selected approaches to unit segmentation**
  - Rule-based unit segmentation using parse trees on essays (Persing and Ng, 2016)
  - Sequence labeling using diverse features on essays (Stab, 2017)
  - Neural argument mining using diverse features on essays (Eger et al., 2017)
  - Cross-genre unit segmentation with various models and features (Ajjour et al., 2017)
  - Contextualized word embeddings with attention on essays (Spliethöver et al., 2019)

Rule-based unit segmentation (Persing and Ng, 2016)

- **Task**
  - Given the sentences of a text, identify all argumentative units.
- **Approach**
  - Get the constituency parse tree of a sentence.
  - Use hand-crafted rules to identify each unit‘s left and right boundary.
- **Data**
  - Corpus with 90 persuasive student essays (Stab and Gurevych, 2014a)
- **Results**
  - 92.1% exact matches of units
  - 98.4% with overlap of ≥ 50%

#	Rules for left-boundaries
1	Exactly where *S* node begins.
2	After initial explicit connective, or if followed by a comma, after comma.
3	After *n*-th comma that is a direct child of *S* node.
4	After *n*-th comma.

#	Rules for right-boundaries
1	At the end of *S* node, or if *S* ends in punctuation, immediately before.
2	If *S* node ends in *SBAR* node, before the n-th shallowest *SBAR*.
3	If *S* node ends in *PP* node, before the *n*-th shallowest *PP*.

Cross-genre unit segmentation (Ajjour et al., 2017)

- **Task**
  - Given a text, classify each token as belonging to an argumentative unit or not.
  - Learn segmentation on one genre, apply on (potentially) another.

- **Research questions**
  1. What features are most effective in unit segmentation?
  2. What model is best to capture relevant context of a token?
  3. To what extent do features and models generalize across genres?

- **Systematic comparison of approaches**
  - Three corpora with texts from different genres
    Basically, all corpora that allowed studying unit segmentation at that time
  - Three machine learning models capturing different context
    The models partly approximate existing approaches, partly realize new ideas.
  - Four feature types capturing different linguistic layers
    The feature types approximate main ideas from previous approaches.

Cross-domain unit segmentation: Corpora

- **Token-level BIO format**
  - Unit segmentation is usually modeled as a BIO tagging task.
  - Each token is beginning (B), inside (I), or outside (O) of an argumentative unit.

> *” If you wanna hear my view I think that the death penalty should be abolished .“*
> O O O O O O O O O B I I I I I O

- **Three corpora of different genres**
  - Essays. 402 persuasive student essays, mean 360 tokens (Stab, 2017)
  - News editorials. 300 news editorials, mean 958 tokens (Al-Khatib et al., 2016)
  - Web discourse. 340 comments etc., mean 253 tokens (Habernal and Gurevych, 2015)

- **Corpus preparations**
  - Annotations boiled down to: argumentative or not
  - Represented in BIO format
  - Original train/test splits used

Corpus	B	I	O
Essays	6 089	94 411	44,022
News editorials	14 234	251 381	21 849
Web discourse	1 129	40 042	44 814

Cross-genre unit segmentation: Approaches

- **What context of a token is important?**

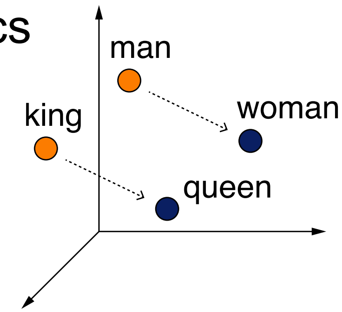
  > ” *If you wanna hear my view I think that* ***the*** *death penalty should be abolished .*“

- **Machine learning models** (details on Bi-LSTM below)
  - SVM. Linear support vector machine that classifies each token independently
  - CRF. Linear-chain conditional random field that classifies each token in the context of its $k = 5$ surrounding tokens
  - Bi-LSTM. Neural network where Bi-LSTMs capture the entire text as context

- **Token-level feature types**
  - Structural. Indicators whether the token is at the start, inside, or at the end of a sentence, clause, or phrase respectively
  - Syntactic. Part-of-speech tag of the token
  - Semantic. The token‘s text (for SVM, CRF) or its embedding (for Bi-LSTM)
  - Pragmatic. Indicators whether the token is before, beginning, inside, end, or after a discourse marker

Background: Word embedding

- **Word embedding** (aka word vector)
  - A real-valued vector that represents the distributional semantics of a particular word in a high-dimensional space



$$\textit{king} \rightarrow \mathbf{v}_{\text{king}} = (0.13, 0.02, 0.1, 0.4, \ldots, 0.22)$$

  - Words that occur in similar contexts have similar embeddings.
    In other words, similarity can be observed even when different words are used.

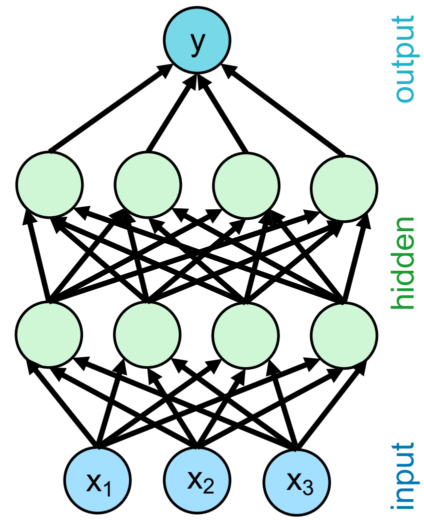
- **Word embedding model**
  - A function that maps each known word to its embedding.
  - Derived from a language model, trained on a (usually huge) corpus

The monarchy is ruled by the _________.

  - Several embedding libraries and pretrained models can be found on the web.
    Libraries: Glove, word2vec, Fasttext, Flair, Bert, ...; **models:** GoogleNews-vectors, ConceptNet Numberbatch, …
  - Many embedding models can also be fine-tuned on a given task.

Background: Neural network

- **Neural network in a nutshell**
  - A network of layers of units that takes a set of input values and computes one or more output values.
  - Used for classification or regression in machine learning
    Deep learning refers to neural networks with multiple hidden layers.
  - Units. Compute non-linear weighted sums of input values
    An activation function, such as *tanh,* is applied to weights learned during training.
  - Layers. Multiple layers allow learning complex functions
    These self-learned functions replace the idea of hand-crafted features.
  - Feed-forward networks. No cycles, fully connected layers (as depicted)

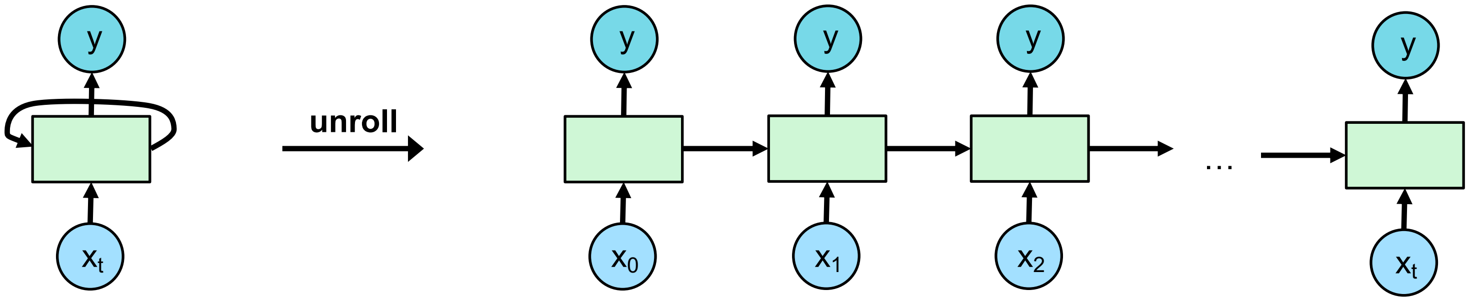


- **Neural networks in NLP**
  - Input tokens are represented in form of word embeddings.
    Other features can still be encoded as *one-hot vectors.*
  - Often, *recurrent neural networks* are used to capture sequential information.

- **Notice**
  - In this course, neural network concepts are detailed only as far as needed.
    For a technical background on neural networks, refer to a machine learning course.

Background: Recurrent neural network based on Jurafsky and Martin (2022)

- **Recurrent neural network (RNN)**
  - A neural network with cycles in its connections, i.e., the value of a unit depends on earlier outputs as an input.



  - A text is processed by presenting one token at a time to the network.
  - The layer from step *i* serves as memory (or context) for decisions in step *j* > *i*.

> ” *If you wanna hear my view I think that* ***the*** *death penalty should be abolished .*“

- **Limitations of simple RNNs**
  - Unidirectionality. Only past input considered, not future input
  - Limited memory. Long-term dependencies hard to learn

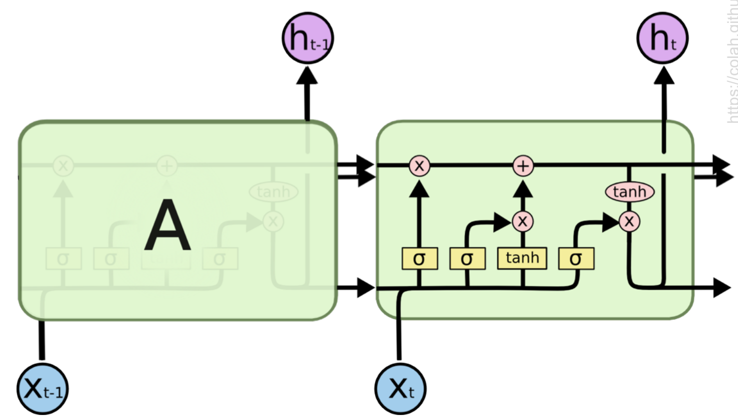
Background: Bi-LSTM neural network

- **Bidirectional RNN**
  - Two RNNs, one processing a text from start to end, the other vice versa

> *” If you wanna hear my view I think that **the** death penalty should be abolished .“*

  - The outputs of the two RNNs are combined into a single representation.
  - By this, an entire input text can be considered as the context of a token.

- **Long short-term memory (LSTM)**
  - Explicit context management in two parts
  - Addition of a context layer to a hidden layer
  - Specialized units that use gates to learn to decide what to forget and what to add for future decisions.



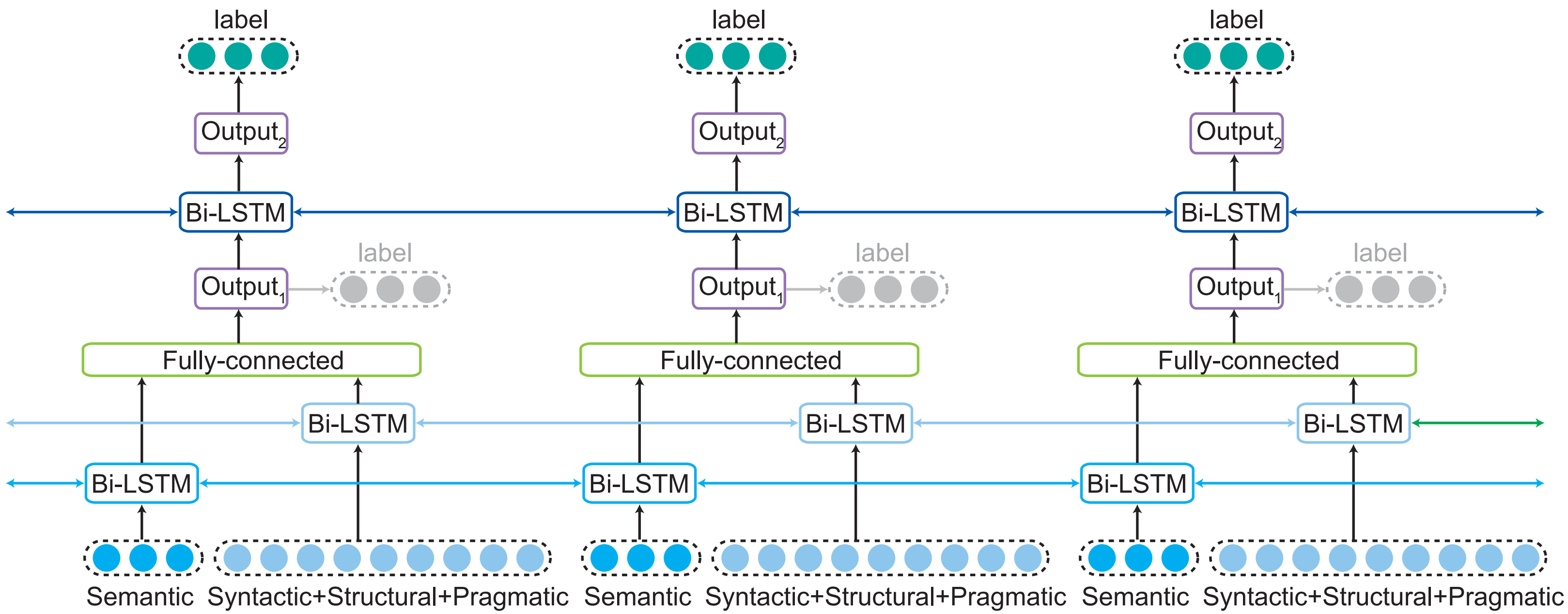
two Bi-LSTM units

- **Bi-LSTM neural network**
  - A bidirectional RNN with LSTM

    Multiple Bi-LSTMs (as well as other neural networks) can easily be stacked.

Cross-genre unit segmentation: Bi-LSTM approach

- **Bi-LSTM unit segmentation approach**



Architecture illustration for three consecutive tokens

- The first Bi-LSTM layers encode semantic features as word embeddings, others as one-hot vectors.
- Another Bi-LSTM layer models dependencies between consecutive tokens.
- Output layers predict confidence values for the possible labels (B, I, O).

Cross-genre unit segmentation: Results

- **Token-level macro F_1-score**
  - All combinations of training and test genre for each approach
    Here, only for *all* features; for details on the features, see the paper of Ajjour et al. (2017).

Approach	Test on essays			Test on news editorials			Test on web discourse		
	Essay	News	Web d.	Essay	News	Web d.	Essay	News	Web d.
SVM	61.4	50.9	31.3	58.8	79.9	**22.6**	**39.1**	37.4	42.8
CRF	79.2	52.5	21.7	**69.8**	82.0	8.0	37.1	**37.6**	37.7
Bi-LSTM	**88.5**	**57.1**	**37.0**	60.7	**84.1**	20.9	20.9	36.6	**54.5**

- **Analysis**
  - 88.5 is significantly better at $p < 0.001$ than best result before (86.7). (Stab, 2017)
  - Bi-LSTM is best, but the others are sometimes better across genres.

  - Semantic features best in-genre (e.g., 87.9 on essays)
  - Structural features most genre-robust (e.g., 35.5–39.5 on web discourse)
  - In general, cross-genre effectiveness limited

Unit segmentation: Discussion

- **Effective unit segmentation**
  - Diverse approaches to unit segmentation may be considered.
  - High effectiveness appears to be possible (only) in narrow explicit genres.
  - The context of a token is critical to assess the token‘s argumentativeness.

- **Definition of argumentative units**
  - The exact difference to syntactic and discourse units remains to be studied.
  - Depending on the genre, units can span anything from clauses to paragraphs.
  - To some extent, unit segmentation is task-specific.

- **Knowledge required for unit segmentation**
  - It is debatable whether unit segmentation should be tackled first.
  - At this point, no knowledge is given about what is argued about.
  - Joint mining approaches may be preferable in some cases. (Eger et al., 2017)

Next section: Unit type classification

I. Introduction to computational argumentation

II. Basics of natural language processing

III. Basics of argumentation

IV. Argument acquisition

V. Argument mining

a) Introduction
b) Argument models
c) Argumentation filtering
d) Unit segmentation
e) Unit type classification
f) Relation identification
g) Conclusion

VI. Argument assessment

VII. Argument generation

VIII. Applications of computational argumentation

IX. Conclusion

What is unit type classification?

- **Unit type classification**
  - The assignment of a class to each argumentative unit from a predefined set of classes, in terms of roles within an argument, evidence types, ...
  - Input. A set of argumentative units, often ordered and grouped by input text
  - Output. Each unit with assigned type

Conclusion

" If you wanna hear my view, I think that the EU should allow rescue boats in the Mediterranean Sea. Many innocent refugees will die if there are no rescue boats.

Premise

Nothing justifies to endanger the life of innocent people."

Premise

- **How to model that computationally?**
  - Supervised text classification of each unit, either feature-based or neural
  - Some approaches tackle unit types as part of relation classification. (more below)

Example: Unit type classification of essay units

- **How good are humans in unit type classification?**
  - Given the following essay units, identify their type (conclusion vs. premise).

Conclusion

” *Living and studying overseas is an irreplaceable experience when it comes to learn standing on your own feet*. *One who is living overseas will of course* Premise *struggle with loneliness, living away from family and friends* but *those difficulties will turn into valuable experiences in the following steps of life*. Premise *Moreover, the one will learn living without depending on anyone else*. Premise

example from Stab and Gurevych (2014a)

- **What makes unit type classification challenging?**
  - Unit types may be issue-dependent, e.g., whether a unit is evidence.
  - Positional information is not always as helpful as for essays.
  - Some types encode structural information, others semantics or pragmatics.

Overview of unit types and their classification

- **Unit types across corpora**
  - Unit types may indicate roles, claim and evidence types, or similar.
    Also, different labels are found in the literature for more or less identical concepts.
  - Unit type schemes are model-specific rather than genre-specific.

- **Selected unit type schemes**
  - Argument roles. AAE, WebDiscourse
    (Stab and Gurevych, 2014a; Habernal and Gurevych, 2015)
  - Claim/evidence. IBM Debater, Webis-16-Editorials
    (Rinott et al., 2015; Al-Khatib et al., 2016)

claim premise none
major claim

premise
rebuttal backing
claim pathos

claim study
anecdotal expert

assumption statistics
anecdote other testimony
common ground

- **Selected approaches to unit type classification**
  - Supervised classification with rich linguistic features (Stab and Gurevych, 2014a; Habernal and Gurevych, 2015; Rinott et al., 2015; Persing and Ng, 2016; Al-Khatib et al., 2017)
  - Unit-level sequence labeling with rich linguistic features (Habernal and Gurevych, 2017)
  - Sequence kernels based on words and part-of-speech (Rooney et al., 2012)
  - Tree kernels based on syntactic parse trees (Liga, 2019)

Supervised classification of evidence types (Al-Khatib et al., 2017)

- **Task**
  - Given editorial units, classify each as being testimony, statistics, anecdote, or none.
- **Approach**
  - Linear SVM on four feature types
  - Lexical. Word *n*-grams
  - Style. Character *n*-grams, length, position, …
  - Syntactic. Part-of-speech *n*-grams
  - Semantic. Entity types, sentiment, …
- **Data**
  - 14k units from 300 editorials (Al-Khatib et al., 2016)
- **Results**
  - Reasonable micro F_1-score, but notable differences across classes

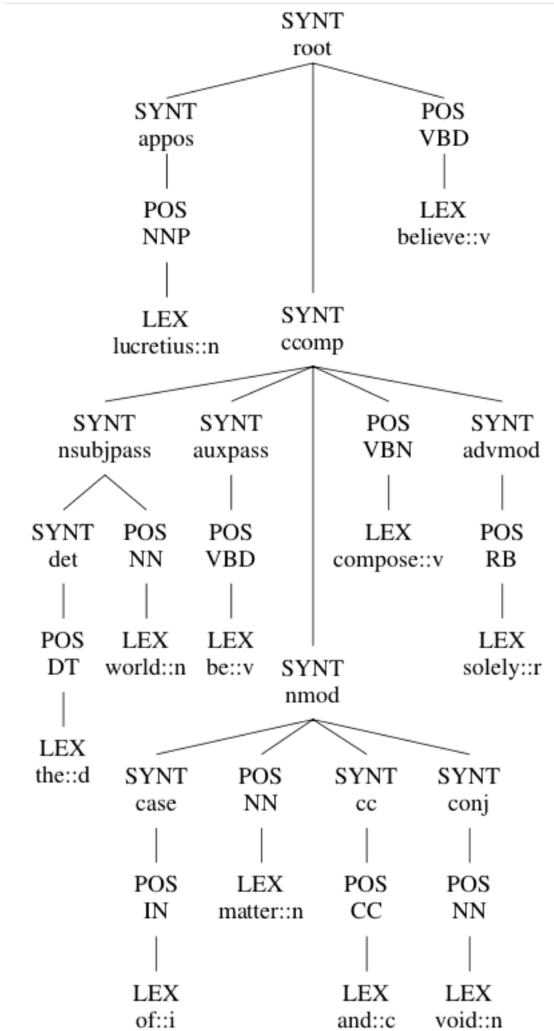
Features	F_1-score
Lexical	0.73
Style	0.70
Syntactic	0.71
Semantic	0.67
All features	**0.77**
Majority baseline	0.56

Class	P	R	F_1
Testimony	0.69	0.40	0.50
Statistics	0.63	0.55	0.59
Anecdote	0.55	0.47	0.51
None	0.84	0.90	0.87

Tree kernels for evidence types (Liga, 2019)

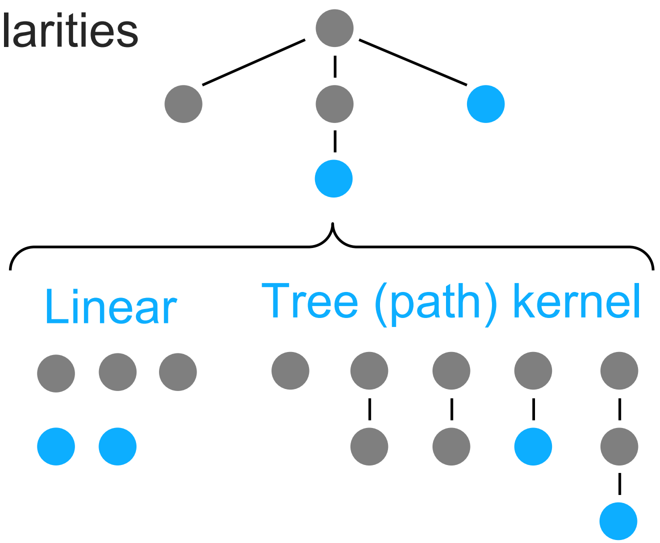
- **Task**
  - Given the argumentative units of a text, classify which of two evidence types each unit has.
- **Research question**
  - Can evidence types be distinguished based on the syntactic structure of a unit?
- **Data**
  - 569 expert/study units from Wikipedia articles (Rinott et al., 2015)
  - 653 testimony/statistics units from news editorials (Al-Khatib et al., 2016)
- **Approach in a nutshell**
  - Tree kernel classification based on a constituency parse tree-like representation of a unit (details below)

"Lucretius believed the world was composed solely of matter and void."



Background: Kernel methods (more in lecture part VI)

- **Kernel methods in machine learning**
  - Kernel methods classify instances by comparing them to known instances.
  - Strong when good features are unknown and data is limited
    Often used for structure input data, such as trees.
- **Kernel method in a nutshell**
  - Kernel. Represents an instance in a task-specific implicit feature space
    Different kernels can be combined mathematically.
  - Similarity function. Quantifies the similarity of any two kernels
  - Classifier. Distinguishes classes based on similarities
    A typical kernel-based classifier is the support vector machine (SVM).
- **Selected kernels for structured data**
  - Linear kernels capture distributions only
    Basically, the correspondent of standard feature vectors.
  - Subsequence kernels for sequential structure
    (Mooney and Bunescu, 2006)
  - Tree kernels for hierarchical structure
    (Collins and Duffy, 2001)



Tree kernels for evidence types: Approach and results

- **Approach**
  - Linear kernel based on the TF-IDF vector of a unit
  - Tree kernel based on the syntax tree of the unit
  - Combined kernel based on the linear and the tree kernel

- **Unit-level macro F_1-score**
  - SVMs on all combinations of training and test genre for each approach

Approach	Test on Wikipedia articles		Test on news editorials	
	Wikipedia	News editorials	Wikipedia	News editorials
Linear kernel (TF-IDF)	0.71	0.72	0.74	0.91
Tree kernel (parse tree)	**0.73**	0.75	0.82	0.87
Combined kernel	0.72	**0.76**	**0.84**	**0.92**

- **Analysis**
  - Tree kernel mostly better than linear kernel; combined kernel best
  - Training on news editorials better on both test sets

Unit type classification: Discussion

- **Unit type classification**
  - Unit type classification is a fairly standard text classification task.
  - Few existing approaches are really argumentation-specific.
  - Often, segmentation and classification are done jointly, or unit classification is done on the sentence level only.

- **Effectiveness of unit type classification**
  - Effectiveness often rather high; on explicit genres, such as essays (F_1 0.87), as well as on more subtle genres, such as news editorials (F_1 0.77).
    (Stab, 2017; Al-Khatib et al., 2017)
  - Still, minority unit types may be hard to classify accurately.

- **Unit types as roles?**
  - Conceptually, classifying the argumentative *role* of a unit is questionable, because one unit may have different roles in different arguments.
  - Still, role classification works well in narrow genres, such as essays. Why?
    Stab (2017) distinguished major claims, claims, premises, and none.

Next section: Relation identification

What is relation identification?

- **Relation identification**
  - The mining of argumentative relations between pairs of argumentative units and the classification of their types, usually as *support* or *attack*
  - Input. A set of argumentative units in a text, possibly with assigned unit type
  - Output. All mined argumentative relations, with their type

Conclusion

> *" If you wanna hear my view, I think that the EU should allow rescue boats in the Mediterranean Sea. Many innocent refugees will die if there are no rescue boats. Nothing justifies to endanger the life of innocent people."*

support

Premise

support

Premise

- **How to model that computationally?**
  - Individual classification of candidate unit pairs
  - Identification of the most likely graph induced by all units and relations

  ... among other ways (more below)

Example: Relation identification of essay units

- **How good are humans in relation identification?**
  - Given the following essay units, mine the relations and classify their types.

> " *Living and studying overseas is an irreplaceable experience when it comes to learn standing on your own feet*. *One who is living overseas will of course struggle with loneliness, living away from family and friends* but *those difficulties will turn into valuable experiences in the following steps of life*. Moreover, *the one will learn living without depending on anyone else*.

attack
support
attack

example from Stab and Gurevych (2014a)

- **What makes relation identification challenging?**
  - Technically, two tasks need to be solved: mining and type classification.
  - In some genres, related units may be far away from each other.
  - Subtle argumentation leaves relations implicit on purpose.

Overview of relations and their identification

- **Argumentative relations across genres**
  - The idea of support and attack is genre-independent.
  - Some argument models consider different relation sub-types.

- **Relations in selected corpora**
  - Essay-specific model. Support and attack by premises (Stab and Gurevych, 2014a)
  - Freeman‘s model. Linked/Convergent support, example, rebuttal, undercutter (Peldsdzus and Stede, 2013)
  - Walton‘s model. Inference relations of argument schemes (Lawrence and Reed, 2017)

- **Selected approaches to relation identification**
  - Maximum spanning tree on classified roles and functions (Peldszus and Stede, 2015)
  - Supervised classification based on topic and discourse (Nguyen and Litman, 2016)
  - Topic modeling based on inferential topic pairs (Lawrence and Reed, 2017)
  - Bi-LSTM neural network based on raw input (Cocarascu and Toni, 2017)
  - Structured SVMs and RNNs based on graph structure (Niculae et al., 2017)

Relation identification using topic modeling (Lawrence and Reed, 2017)

- **Task**
  - Given an ADU pair *(P, C)*, decide whether they are in argumentative relation.
- **Approach**
  - Crawl related web texts based on top word 1- and 2-grams in dataset.
  - Extract inferential sentence pairs from web texts using high-precision indicators.
  - Use LDA topic modeling on sentences to get premise/conclusion topic probabilities.
  - Relate pair if probability is above average.

Indicator	P	R
P *therefore* C	.95	.0004
C *because* P	.91	.0031
P *consequently* C	.82	.0001
P *hence* C	.76	.0001
P *accordingly* C	.74	.0002

- **Data**
  - 327 ADUs with 128 relations from transcripts of a political radio program
- **Results**
  - Max F_1-score at high recall, but low precision

Class	P	R	F_1
Random	0.50	0.50	0.50
Approach	0.64	0.83	0.72

MST relation identification (Peldszus and Stede, 2015)

- **Task**
  - Given the segmented ADUs of a text, mine relations between the ADUs and classify them as support or attack.

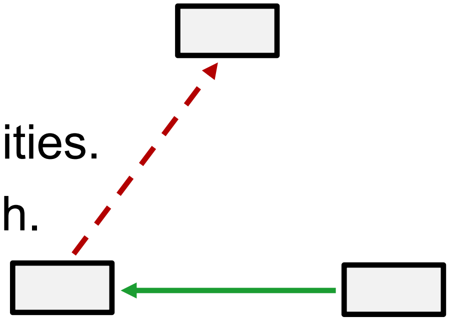
    Peldszus and Stede (2015) study further tasks left out here for simplicity.

- **Research question**
  - Does information about unit types and other argumentative relations in a text help to mine and classify relations?

- **Data**
  - Arg-microtexts. 112 texts with 576 ADUs, annotated for Freeman‘s model
  - The relations are simplified to (single) support and attack.

    290 support, 174 attack, and 2000 ADU pairs w/o relation.

- **Approach**
  - Supervised classifiers to obtain role and function probabilities.
  - Weighted probability aggregation to obtain evidence graph.
  - Maximum spanning tree (MST) to obtain relations.



MST relation identification: Example relations

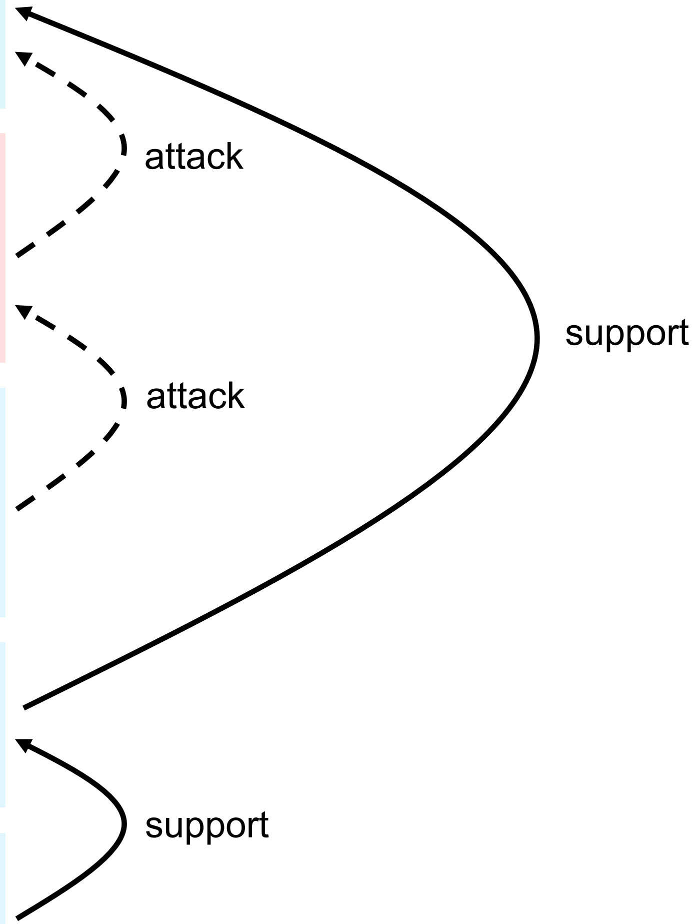
Health insurance companies should naturally cover alternative medical treatments.

Not all practices and approaches that are lumped together under this term may have been proven in clinical trials.

Yet it's precisely their positive effect when accompanying conventional 'western' medical therapies that's been demonstrated as beneficial.

Besides many general practitioners offer such counselling and treatments in parallel anyway

and who would want to question their broad expertise?



MST relation identification: Classifiers and aggregation

- **Supervised classifiers**
  - Role. Predict probabilities of an ADU being proponent (p_p) and thesis (p_t)
  - Function. Predict probability of an ADU being a support (p_s)
  - Relation. Predict probability of an ADU pair being in relation ($p_r^{(i,j)}$)

    A log-loss model with stochastic gradient is used in each case in the experiments.

- **Employed features**
  - Content. Lemma *n*-grams
  - Style. POS tags, main verb morphology, discourse connectives, …
  - Structure. Length of ADU, position in text, …
  - ADU pair. Distance and order of the candidate ADU pair

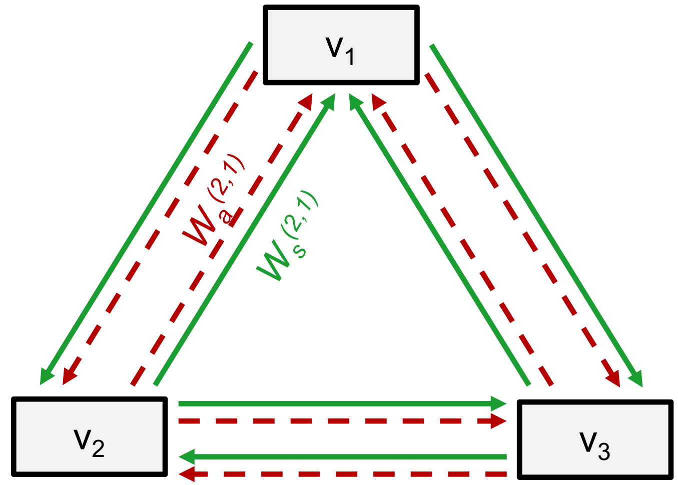
- **Weighted probability aggregation**
  - Single scores $p^{(i,j)}$ for each ADU pair can be derived from the probabilities.

    Details left out here for simplicity
  - A weighted pair score can be learned in training.

$$w^{(i,j)} = \frac{w_p \cdot p_p^{(i,j)} + w_t \cdot p_t^{(i,j)} + w_s \cdot p_s^{(i,j)} + w_r \cdot p_r^{(i,j)}}{\sum_k w_k}$$

MST relation identification: Evidence graph and MST

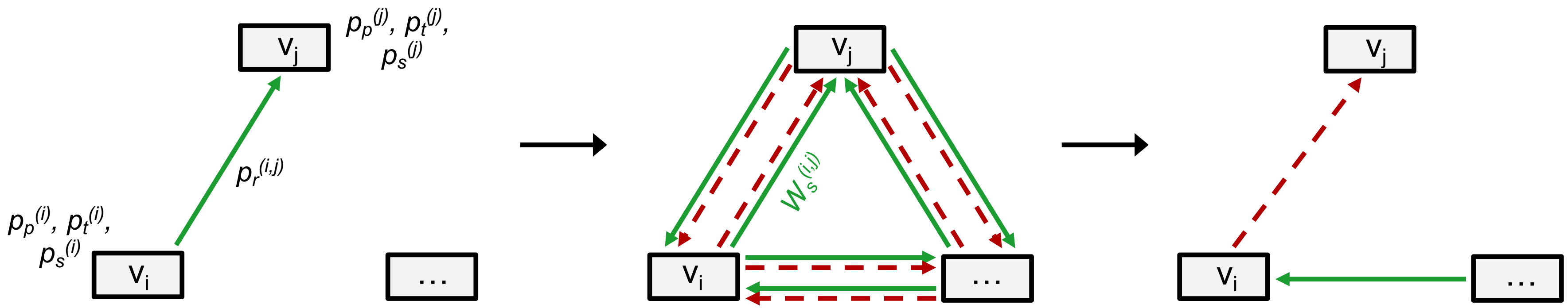
- **Evidence graph**
  - A weighted directed graph $G = (V, E)$
  - Nodes. Each node v in V represents an ADU.
  - Support edges. Any pair of nodes v_i, v_j is connected with an edge e_s.
  - Attack edges. Any pair of nodes v_i, v_j is connected with an edge e_a.
  - Weights. Each e is labeled with a weighted pair score $w^{(i,j)}$ as defined above.



- **Maximum spanning tree (MST)**
  - A sub-graph G^* of a weighted graph $G = (V, E)$ whose edges E connect all nodes V and that has maximum weight.
  - Finding an MST can be solved efficiently for directed and undirected graphs.
    This should be known from basic algorithm classes.
  - Chu-Liu-Edmonds algorithm. Finds MSTs of directed graphs in $O(E + V \log V)$.
    (Chu and Liu, 1965; Edmonds, 1967)

MST relation identification: Approach and baselines

- **Approach**
  - Apply classifiers and weighted aggregation to build evidence graph.
  - Apply Chu-Liu-Edmonds algorithm to obtain MST.



- **Baselines**
  - Classifiers. Determine whether one ADU supports or attacks another, or not.
  - MST parser. An off-the-shelf discourse parser that uses structured learning to build the MST based on the discourse structure of a text
  - MST parser + classifiers. The MST parser, using the classifiers‘ outputs as additional features during training

MST-based relation identification: Results

- **Macro F1-score on relation identification**
  - Mining. Source ADU is a premise of target ADU or not
  - Classification. ADU has a supporting or an attacking function

  Jointly trained in 5-fold cross validation, averaged over 10 runs; more results found in Peldszus and Stede (2015)

Approach	Mining	Classification
Classifiers	0.66	0.67
MST parser	0.71	0.49
MST parser + classifiers	**0.72**	0.68
Approach	0.69	**0.71**

- **Analysis**
  - The approach turns out best in classifying relations, but not in mining them.
  - The classifiers also work well with the off-the-shelf MST parser.
  - The MST idea makes sense, if full argumentative structure can be expected.
  - Otherwise, some kind of argument decomposition may be needed before.

Relation identification: Discussion

- **Relation identification**
  - Diverse approaches have been proposed for relation identification.
  - Many works focus on *support*, the default relation from premise to conclusion.
  - Unlike in this lecture part, relations may also be identified for *argument* pairs.

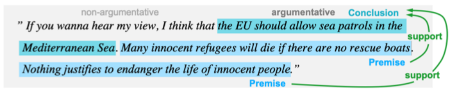
- **Effectiveness of relation identification**
  - Mining relations usually works better than classifying attack vs. support.
  - Semi-reliable for explicit argumentation (Stab, 2017)
  - Unsolved for ”hidden“ argumentation, even hard for humans (Al-Khatib et al., 2017)

- **Difference to stance**
  - Attack/support and pro/con stance classification conceptually overlap.
  - Unlike relations, stance actually refers to the author‘s position on an issue.
  - Still, support/attack can be modeled as pro/con premises with hardly any loss. (Wachsmuth et al., 2017f)

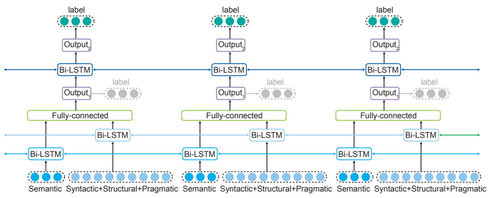
Next section: Conclusion

Conclusion

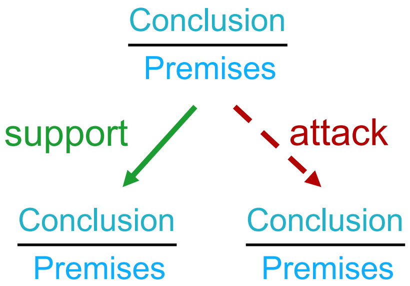
- **Argument mining**
  - Computational identification of argumentative structure
  - May be based on different argument models
  - Segmenting units, classifying types, identifying relations



- **Selected approaches to argument mining**
  - Unit segmentation using rules or Bi-LSTMs
  - Unit type classification using features or tree kernels
  - Relation identification using topic modeling or MSTs



- **Discussion of argument mining**
  - May work pretty reliable within narrow, explicit genres
  - Hard on subtle argumentation, unsolved across genres
  - Simple argument models may allow more robustness



References

- **Ajjour et al. (2017)**. Yamen Ajjour, Wei-Fan Chen, Johannes Kiesel, Henning Wachsmuth, and Benno Stein. Unit Segmentation of Argumentative Texts. In Proceedings of the Fourth Workshop on Argument Mining, pages 118–128, 2017.
- **Al-Khatib et al. (2016).** Khalid Al Khatib, Henning Wachsmuth, Johannes Kiesel, Matthias Hagen, and Benno Stein. A News Editorial Corpus for Mining Argumentation Strategies. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 3433–3443, 2016.
- **Al-Khatib et al. (2017).** Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, and Benno Stein. Patterns of Argumentation Strategies across Topics. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1362–1368, 2017.
- **Amgoud and Ben-Naim (2018).** Leila Amgoud and Jonathan Ben-Naim. Weighted Bipolar Argumentation Graphs: Axioms and Semantics. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, pages 5194–5198, 2018.
- **Bar-Haim et al. (2017).** Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. Stance Classification of Context-Dependent Claims. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pages 251–261, 2017.
- **Baldrige et al. (2007).** Jason Baldridge, Nicholas Asher, and Julie Hunter. 2007. Annotation for and Robust Parsing of Discourse Structure on Unrestricted Texts. *Zeitschrift für Sprachwissenschaft*, 26:213–239.
- **Chu and Liu (1965).** Y. J. Chu and T. H. Liu. 1965. On the Shortest Arborescence of a Directed Graph. Science Sinica, 14:1396–1400.
- **Cocarascu and Toni (2017).** Identifying Attack and Support Argumentative Relations using Deep Learning. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1385–1390, 2017.

References

- **Collins and Duffy (2001).** Michael Collins and Nigel Duffy. 2001. Convolution kernels for natural language. In Advances in Neural Information Processing Systems 14, pages 625–632.
- **Dung (1995):** Phan Minh Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. Artificial Intelligence, 77(2):321–357, 1995.
- **Edmunds (1967).** Jack Edmonds. 1967. Optimum Branchings. Journal of Research of the National Bureau of Standards, 71B:233–240.
- **Eger et al. (2017).** Steffen Eger, Johannes Daxenberger, and Iryna Gurevych. Neural end-to-end learning for computational argumentation mining. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 11–22, 2017.
- **Freeman (2011).** Argument Structure: Representation and Theory. Springer, 2011.
- **Habernal and Gurevych (2015)***.* Ivan Habernal and Iryna Gurevych. Exploiting Debate Portals for Semi-supervised Argumentation Mining in User-generated Web Discourse. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 2127– 2137, 2015.
- **Habernal and Gurevych (2017)***.* Ivan Habernal and Iryna Gurevych. Argumentation mining in user-generated web discourse. Computational Linguistics, 43(1), pages 125–179, 2017.
- **Jurafsky and Martin (2022).** Daniel Jurafsky and James H. Martin (2022). Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics. 3rd edition draft. https://web.stanford.edu/~jurafsky/slp3/
- **Kunz and Rittel (1970).** Werner Kunz and Horst W. J. Rittel. Issues as Elements of Information Systems. Technical Report Working Paper No. 131, Institute of Urban and Regional Development, University of California, Berkeley, 1970

References

- **Lawrence and Reed (2017).** John Lawrence and Chris Reed. Mining Argumentative Structure from Natural Language text using Automatically Generated Premise-Conclusion Topic Models. In Proceedings of the 4th Workshop on Argument Mining, pages 39–48, 2017.
- **Liga (2019).** Davide Liga. Argumentative Evidences Classification and Argument Scheme Detection Using Tree Kernels. In Proceedings of the 6th Workshop on Argument Mining, pages 92–97, 2019.
- **Mooney and Bunescu (2006).** Raymond J. Mooney and Razvan C. Bunescu. 2006. Subsequence kernels for relation extraction. In *Advances in Neural Information Processing Systems 18*, pages 171–178. MIT Press.
- **Nguyen and Litman (2016).** Huy V. Nguyen and Diane J. Litman. Context-aware Argumentative Relation Mining. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pages 1127–1137, 2016.
- **Niculae et al. (2017).** Vlad Niculae, Joonsuk Park, and Claire Cardie. Argument Mining with Structured SVMs and RNNs. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 985–995, 2017.
- **Rinott et al. (2015).** Ruty Rinott, Lena Dankin, Carlos Alzate Perez, M. Mitesh Khapra, Ehud Aharoni, and Noam Slonim. Show Me Your Evidence — An Automatic Method for Context Dependent Evidence Detection. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 440–450, 2015.
- **Peldszus and Stede (2015).** Andreas Peldszus and Manfred Stede. Joint Prediction in MST-style Discourse Parsing for Argumentation Mining. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 938–948, 2015.
- **Persing and Ng (2016).** Isaac Persing and Vincent Ng. End-to-End Argumentation Mining in Student Essays. In Proceedings of NAACL-HLT 2016, pages 1384–1394, 2016.

References

- **Rooney et al. (2019).** Niall Rooney, Hui Wang, and Fiona Browne. Applying kernel methods to argumentation min- ing. In Proceedings of the 25th FLAIRS Conference, pages 272–275, 2012.
- **Spliethöver et al. (2019).** Maximilian Spliethöver, Jonas Klaff, and Hendrik Heuer. Is It Worth the Attention? A Comparative Evaluation of Attention Layers for Argument Unit Segmentation. In Proceedings of the 6th Workshop on Argument Mining, pages 74–82, 2019.
- **Stab (2017).** Christian Stab. Argumentative Writing Support by means of Natural Language Processing, Chapter 5. PhD thesis, TU Darmstadt, 2017.
- **Stede and Schneider (2018).** Manfred Stede and Jodi Schneider. Argumentation Mining. Synthesis Lectures on Human Language Technologies 40, Morgan & Claypool, 2018.
- **Toulmin (1958).** Stephen E. Toulmin. The Uses of Argument. Cambridge University Press, 1958.
- **Wachsmuth et al. (2017f).** Henning Wachsmuth, Giovanni Da San Martino, Dora Kiesel, and Benno Stein. The Impact of Modeling Overall Argumentation with Tree Kernels. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 2369–2379, 2017.
- **Walton et al. (2008).** Douglas Walton, Christopher Reed, and Fabrizio Macagno. Argumentation Schemes. Cambridge University Press, 2008.