

Argument Mining: Optimizing Search and Decision Processes by Means of Large-Scale Unstructured Text Data



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Iryna Gurevych
UKP Lab
Computer Science

What is Argument Mining?

Argument Mining

- Recognize arguments in text automatically
- Assessing the quality of textual arguments
- Based on supervised machine learning

Discourse-level perspective

- Argument components
- Argumentative structures
- Single document of specific type



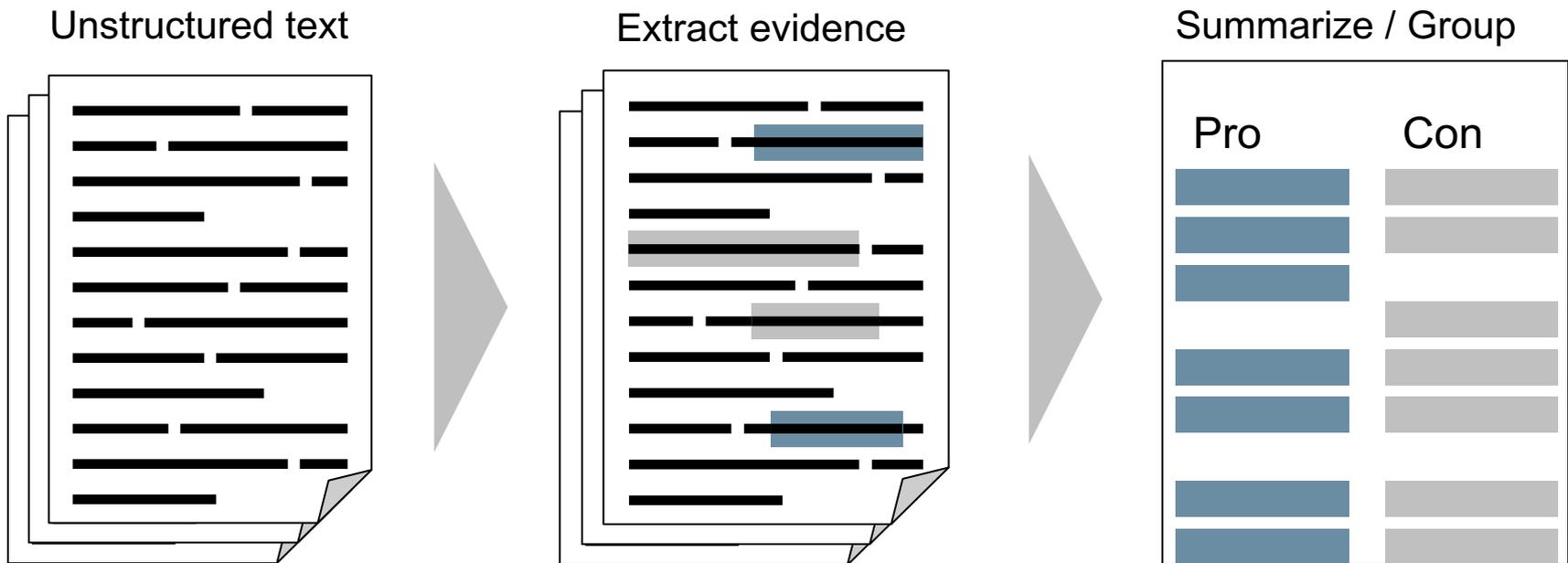
Information-seeking perspective (focus of this talk)

- Arguments relevant to a given topic
- Multiple documents

Goals of Information-Seeking Perspective

Given: a controversial **topic** (e.g. “autonomous cars” or “basic income”)

Extract **pro and con arguments** from different kinds of text



ArgumenText: <http://www.argumenttext.de/showcases/>



Pro/Con List Weights Docs

Found 164 arguments (98 pro; 66 con) in 20 documents (classified 621 sentences in 2.921 ms)

Filter by URL:

- multivu.com (21)
- futurist.com (18)
- gizmodo.com (15)
- blog.cjponyparts.com (13)
- self-drivecar.com (11)
- ideas.time.com (10)
- businessinsider.com (9)
- bgr.com (9)
- thetruthaboutcars.com (7)
- extremetech.com (7)
- hybridcars.com (6)
- computerworld.com (6)
- dailybits.com (6)
- autoworldnews.com (6)
- whogotfunded.com (5)
- mobilenapps.com (5)
- slashgear.com (4)

PRO: Thanks to vehicle-to-vehicle and vehicle-to-infrastructure communication systems, autonomous cars and trucks could significantly reduce traffic congestion and traffic accidents. (0.9771)
<http://www.futurist.com/2013/11/13/greener-future-self-driving-cars/>

PRO: Self-driving vehicles can contribute to reducing infrastructure investments and enrich city life in other ways, such as by reducing emissions, and improving air quality and traffic safety. (0.9709)
<http://www.multivu.com/mnr/64153-volvo-self-driving-cars-unique-swedish-project>

PRO: Autonomous vehicles and a smarter infrastructure will bring us another step closer to even safer traffic and an improved environment. (0.9711)
<http://www.multivu.com/mnr/64153-volvo-self-driving-cars-unique-swedish-project>

PRO: This technology can also improve safety significantly, reduce fuel consumption and congestion. (0.9669)
<http://www.multivu.com/mnr/64153-volvo-self-driving-cars-unique-swedish-project>

CON: Because self-driving cars are powered by computers and computers can inevitably be hacked, there are some serious security concerns here. (0.9592)
<http://gizmodo.com/whats-keeping-self-driving-cars-off-the-road-1450916024>

CON: However, it's important to point out that self-driving cars could also pose some unique safety problems of their own. (0.9545)
<http://blog.cjponyparts.com/2014/08/are-we-road-self-driving-cars-infographic/>

CON: For example, we haven't built self-driving cars to deal with intense conditions like busy city driving and extreme weather. (0.9204)
<http://gizmodo.com/whats-keeping-self-driving-cars-off-the-road-1450916024>

CON: The technology's not quite thereThe most obvious hurdle for self-driving cars is the technology that makes them drive themselves. (0.8659)
<http://gizmodo.com/whats-keeping-self-driving-cars-off-the-road-1450916024>

Research Challenges

- Challenge 1:** Annotating arguments in heterogeneous texts
- Challenge 2:** Creating large amounts of training data
- Challenge 3:** Training models robust enough for different topics

Challenge 1

Annotating arguments in **heterogeneous** texts



Requirements

1. Applicable to information seeking perspective
2. General enough for heterogeneous texts
3. Simple enough for crowdsourcing

Our solution

- **Topic** is some matter of controversy that can be expressed with keywords
- **Argument** is a span of text with evidence supporting or opposing a given topic
- **Three classes sentence-wise:** pro, con, no argument

Annotation Model: Examples

Three classes sentence-wise:
pro, con, no argument

Topic	Sentence	Label
nuclear energy	Nuclear fission is the process that is used in nuclear reactors to produce energy using element called uranium.	?
nuclear energy	The amount of greenhouse gases have decreased by almost half because of the prevalence in the utilization of nuclear power.	
minimum wage	A 2014 study [. . .] found that minimum wage workers are more likely to report poor health and suffer from chronic diseases.	
minimum wage	We should abolish all Federal wage standards and allow states and localities to set their own minimums.	

Annotation Model: Examples

Three classes sentence-wise:
pro, con, no argument

Topic	Sentence	Label
nuclear energy	Nuclear fission is the process that is used in nuclear reactors to produce energy using element called uranium.	no argument
nuclear energy	The amount of greenhouse gases have decreased by almost half because of the prevalence in the utilization of nuclear power.	?
minimum wage	A 2014 study [. . .] found that minimum wage workers are more likely to report poor health and suffer from chronic diseases.	
minimum wage	We should abolish all Federal wage standards and allow states and localities to set their own minimums.	

Annotation Model: Examples

Three classes sentence-wise:
pro, con, no argument

Topic	Sentence	Label
nuclear energy	Nuclear fission is the process that is used in nuclear reactors to produce energy using element called uranium.	no argument
nuclear energy	The amount of greenhouse gases have decreased by almost half because of the prevalence in the utilization of nuclear power.	pro argument
minimum wage	A 2014 study [. . .] found that minimum wage workers are more likely to report poor health and suffer from chronic diseases.	?
minimum wage	We should abolish all Federal wage standards and allow states and localities to set their own minimums.	

Annotation Model: Examples

Three classes sentence-wise:
pro, con, no argument

Topic	Sentence	Label
nuclear energy	Nuclear fission is the process that is used in nuclear reactors to produce energy using element called uranium.	no argument
nuclear energy	The amount of greenhouse gases have decreased by almost half because of the prevalence in the utilization of nuclear power.	pro argument
minimum wage	A 2014 study [. . .] found that minimum wage workers are more likely to report poor health and suffer from chronic diseases.	con argument
minimum wage	We should abolish all Federal wage standards and allow states and localities to set their own minimums.	?

Annotation Model: Examples

Three classes sentence-wise:
pro, con, no argument

Topic	Sentence	Label
nuclear energy	Nuclear fission is the process that is used in nuclear reactors to produce energy using element called uranium.	no argument
nuclear energy	The amount of greenhouse gases have decreased by almost half because of the prevalence in the utilization of nuclear power.	pro argument
minimum wage	A 2014 study [. . .] found that minimum wage workers are more likely to report poor health and suffer from chronic diseases.	con argument
minimum wage	We should abolish all Federal wage standards and allow states and localities to set their own minimums.	no argument

Annotation Model: Expert Annotations

Data

- Heterogeneous text types (news, online discussions, blogs, social media, etc.)
- Eight controversial topics, e.g. “school uniforms”, “gun control”, etc.
- Collected from web searches (query Google for topic)

Annotation Study

- Two expert annotators
- Graduate-level language technology researchers
- Independent annotation of 200 sentences for each topic (1.600 total)

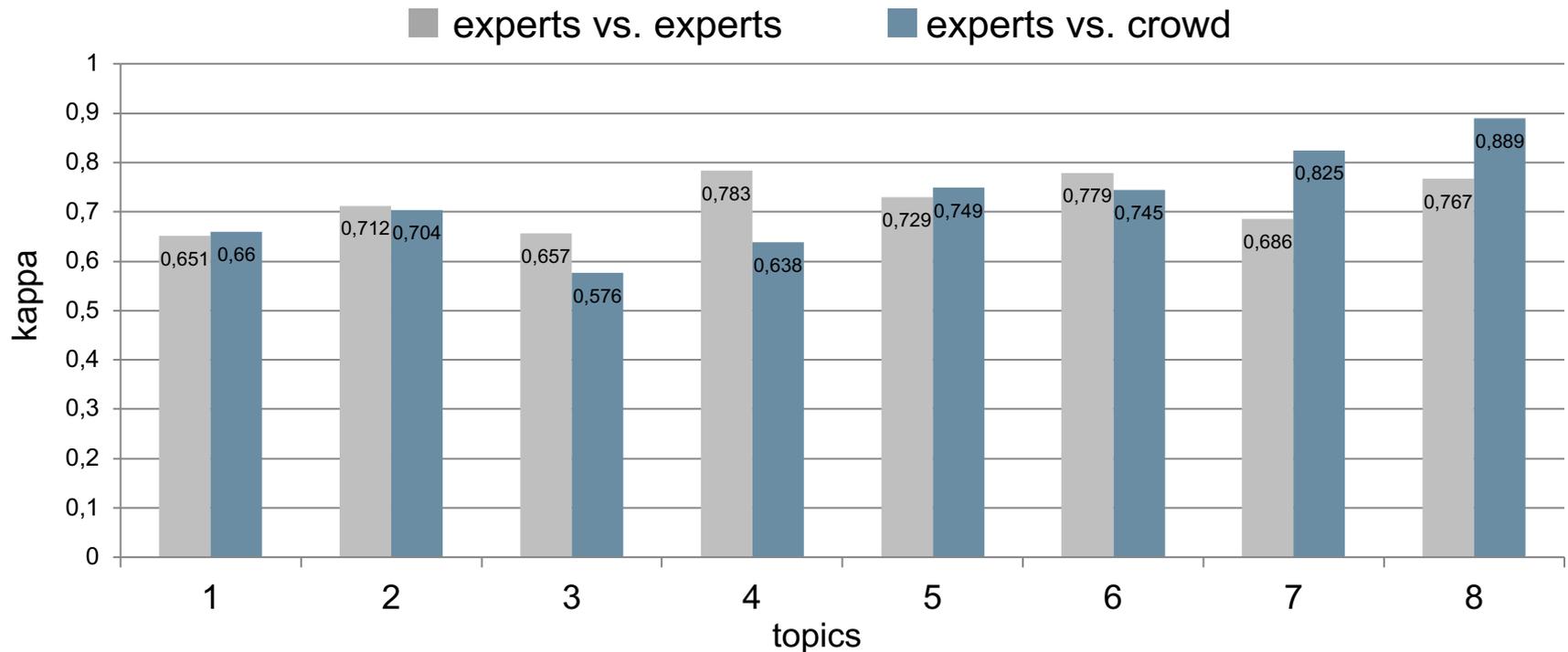
Average agreement over topics

- $\kappa = 0.721$
- **Sufficient agreement:** Annotation model is applicable to heterogeneous texts by expert annotators

Challenge 2

Creating **large** amounts of training data

Comparing Experts to Crowdworkers



Results

- High quality annotations using crowdsourcing
- Crowdworkers achieve kappa = .723 agreement with expert annotations

Statistics of Final Corpus



topic	docs	sentences	no argument	support argument	oppose argument
abortion (AB)	50	3,929	2,427	680	822
cloning (CL)	50	3,039	1,494	706	839
death penalty (DP)	50	3,651	2,083	457	1,111
gun control (GC)	50	3,341	1,889	787	665
marijuana legalization (ML)	50	2,475	1,262	587	626
minimum wage (MW)	50	2,473	1,346	576	551
nuclear energy (NE)	50	3,576	2,118	606	852
school uniforms (SU)	50	3,008	1,734	545	729
total	400	25,492	14,353	4,944	6,195

Table 3: Corpus size and class distribution.

- Annotation process is scalable: **25k+ instances in less than a week**
- Costs: **\$2,774**
- Corpus allows learning a classifier for argument mining **across topics**

Challenge 3

Training a classifier robust enough for **different topics**

Experimental Setup

Experiments

1. Can we improve accuracy by leveraging the topic?
2. Does more training data improve the results?

Evaluation setup

- Task: classify a sentence as “argument” or “no argument” relevant to the topic
- In-domain: train and test on the same topic
- Cross-domain: train on $n-1$ topics and test on left-out-topic

Experiment 1: Models

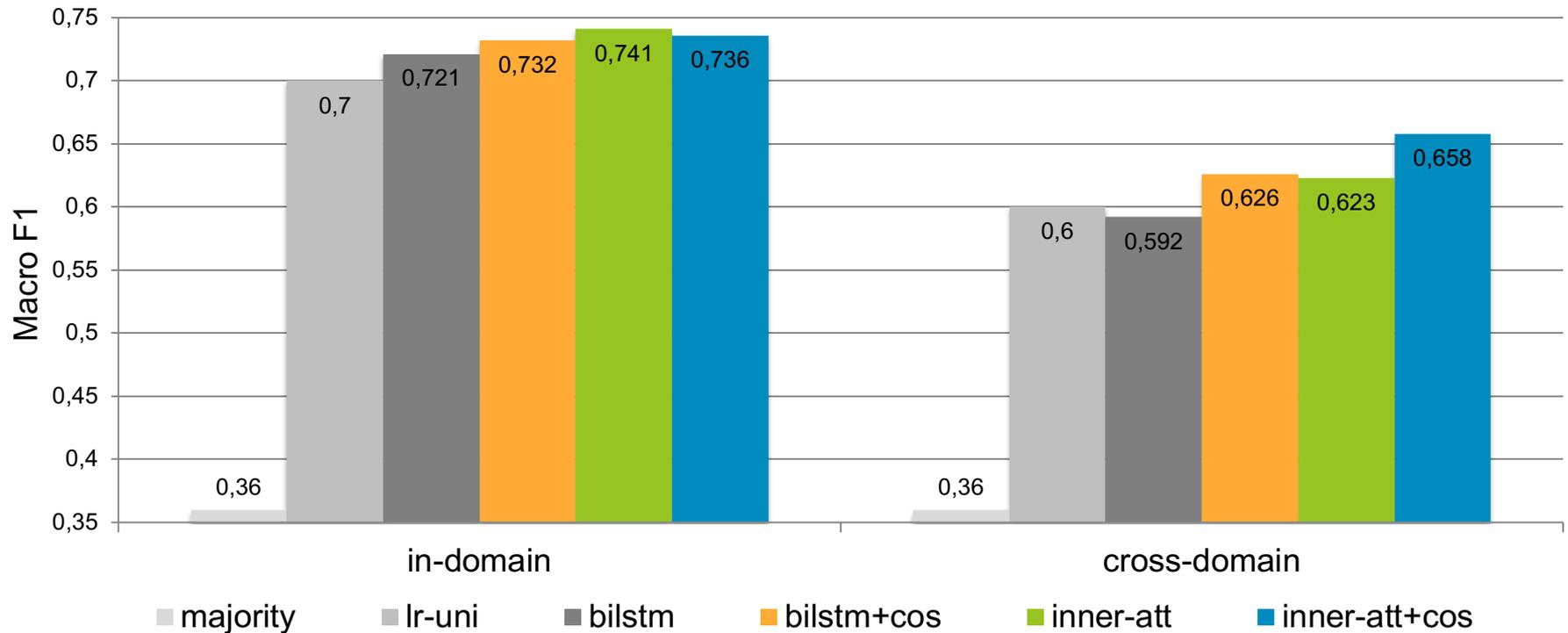
Baselines

- **majority**: classifies each instance as “no argument”
- **lr-uni**: logistic regression with binary unigram features
- **bilstm**: bidirectional long short-term memory network 300d embeddings

Models with topic information

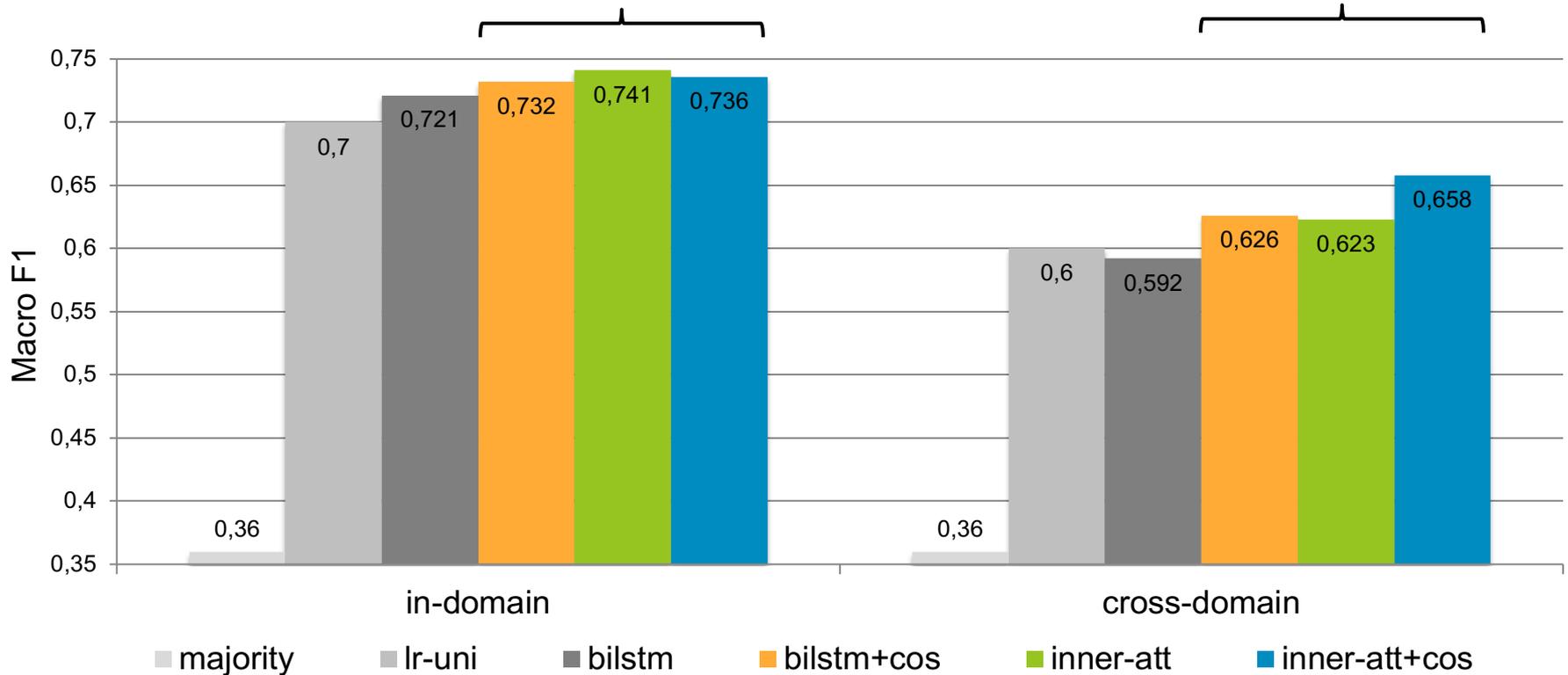
- **bilstm+cos**: bilstm model with topic similarity feature
- **inner-att**: learns weighting of input word with respect to the given topic
- **inner-att+cos**: combines bilstm+cos and inner-att models

Experiment 1: Evaluation



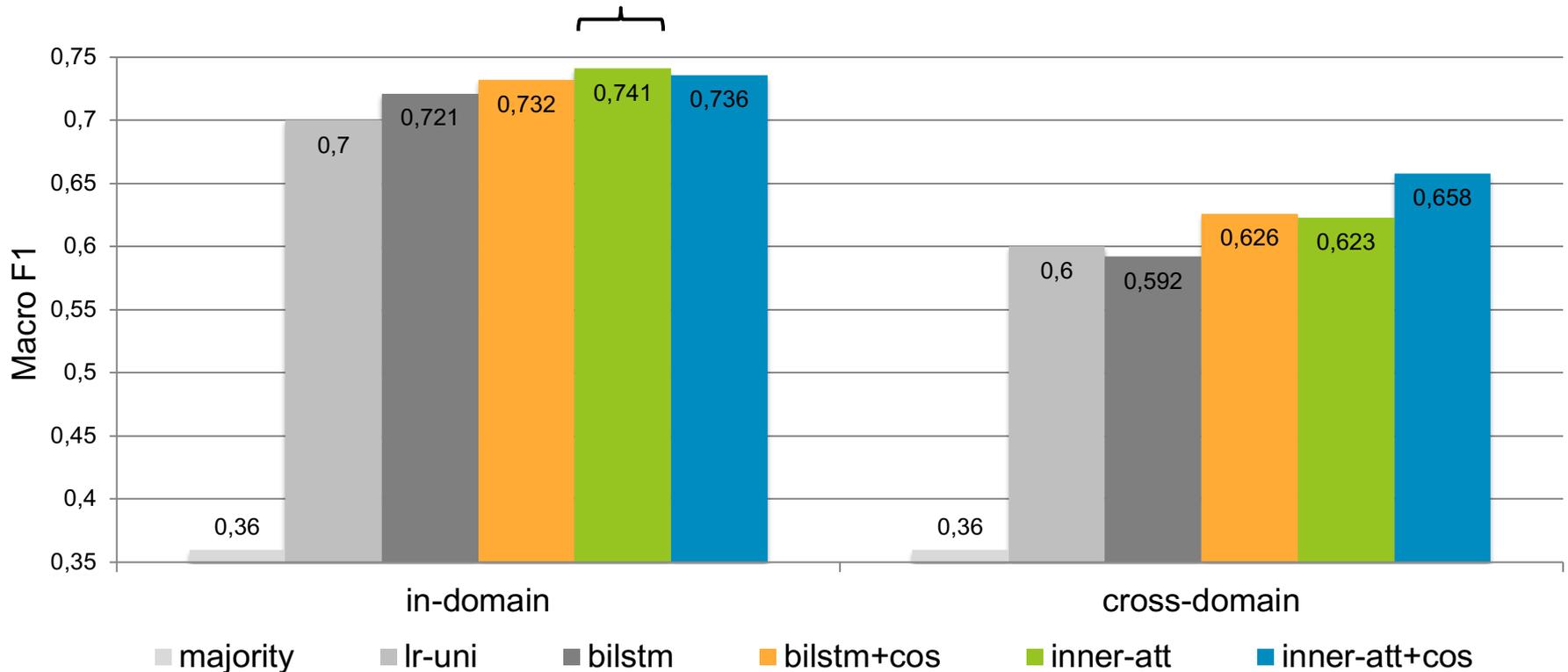
Experiment 1: Evaluation

**Adding topic information improves
baseline results**



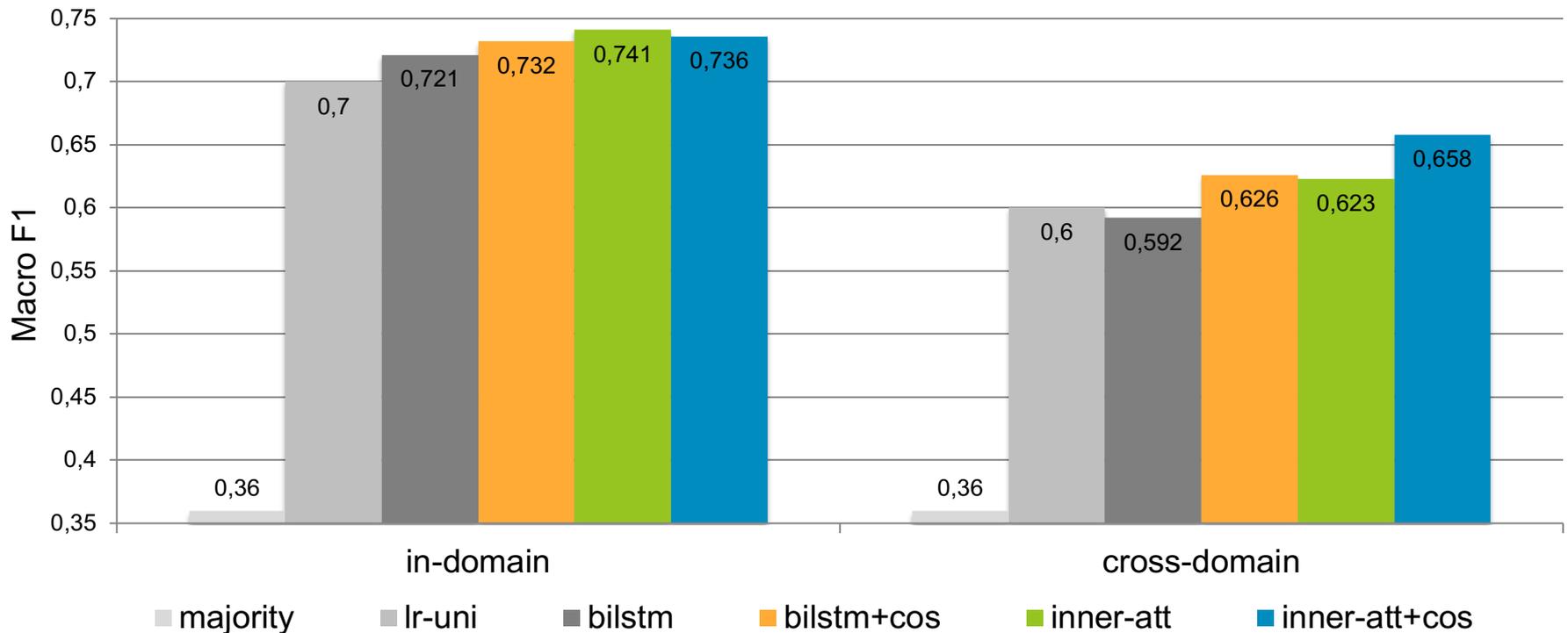
Experiment 1: Evaluation

Inner-att achieves best in-domain results



Experiment 1: Evaluation

**Inner-att+cos generalizes
best to unknown topics**



Experiment 2: Corpus Extension

Does the model generalize better if more topics are in the training data?

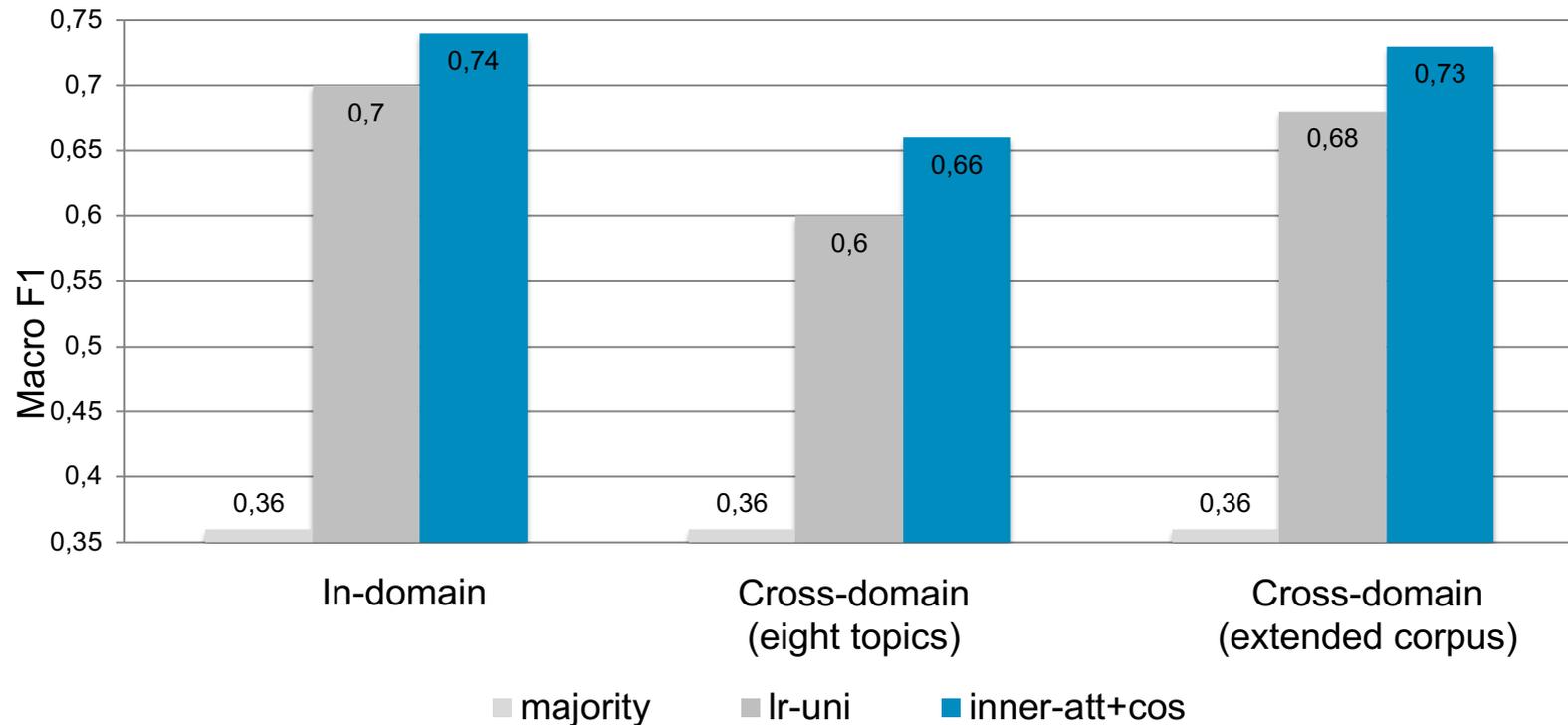
Corpus Extension

- Add additional 41 topics to our training data
 - e.g. “autonomous driving”, “cryptocurrency”, “drones”, “biofuel”, etc.
- Per topic ~600 additional annotated instances

Size of extended corpus

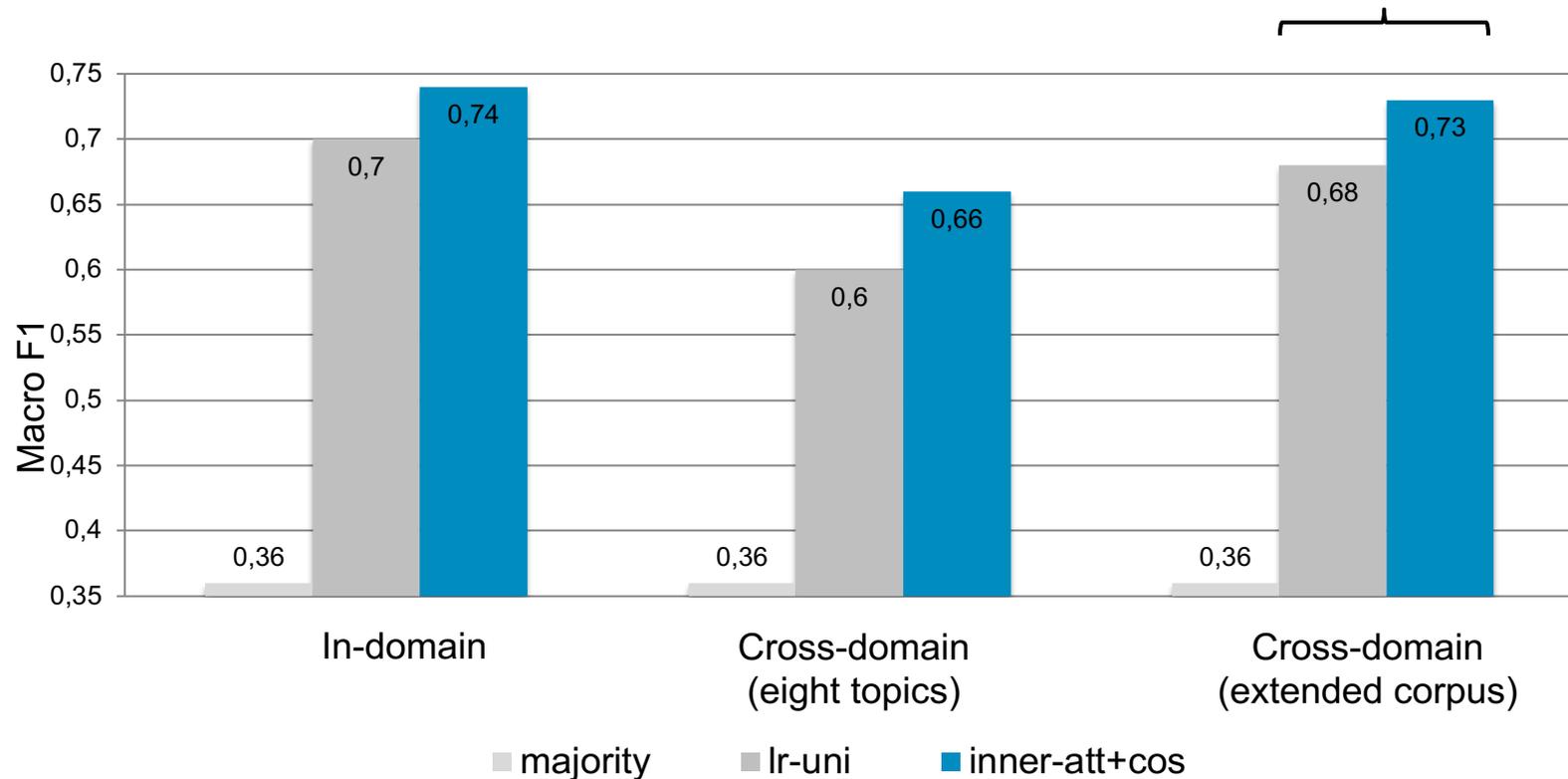
- 49 topics
- 50k+ instances

Results Using Extended Corpus



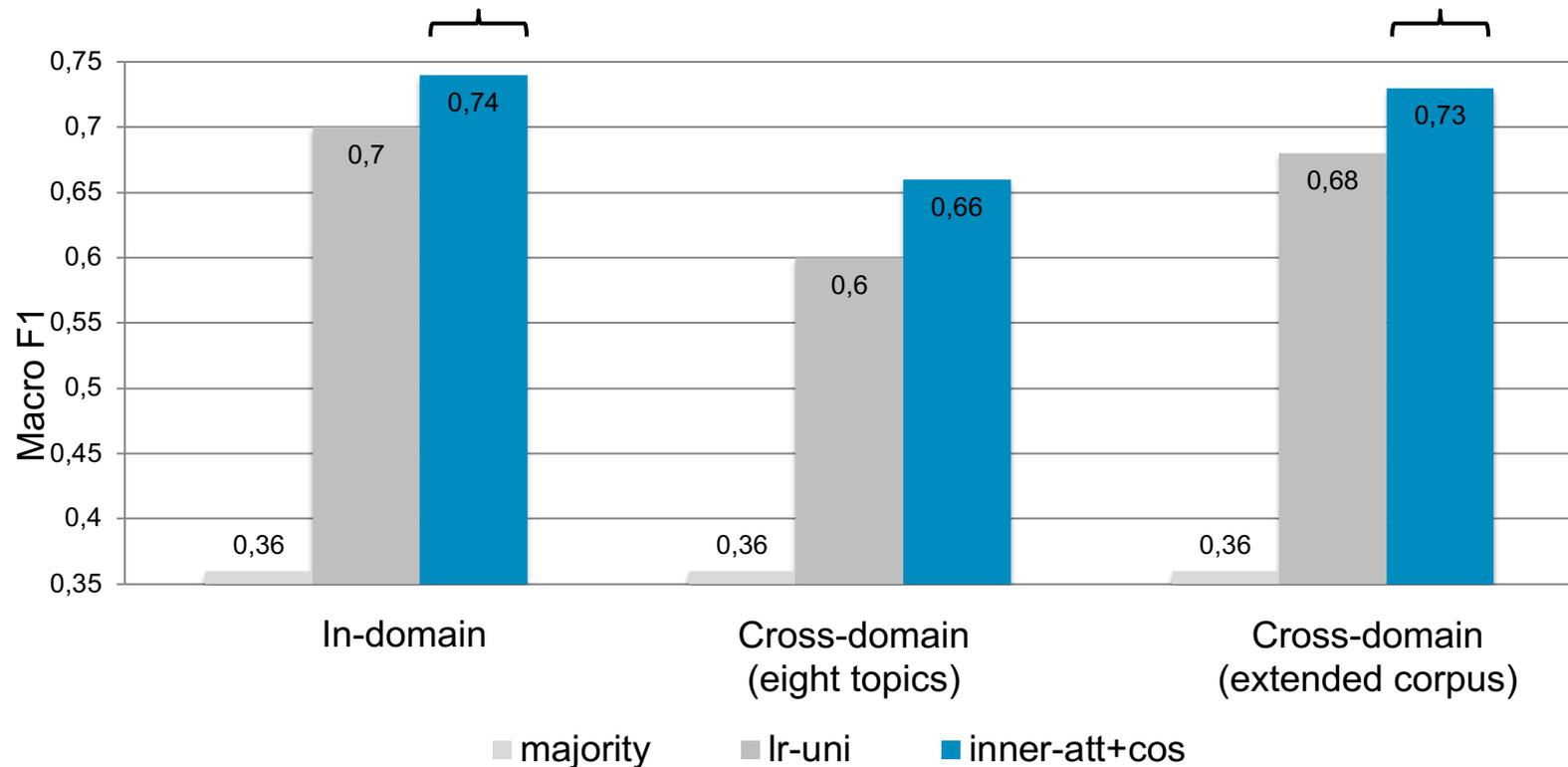
Results Using Extended Corpus

Adding more topics to training data helps



Results Using Extended Corpus

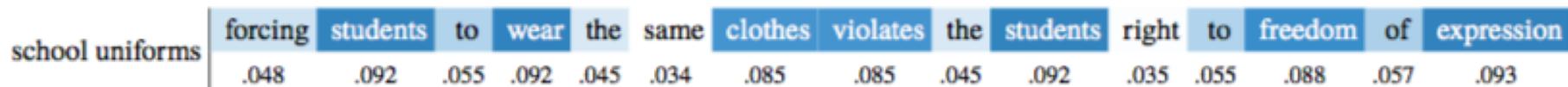
Inner-att+cos achieves almost in-domain results without seeing the topic in test data



What Does the Model Learn?

Visualization of attention weights

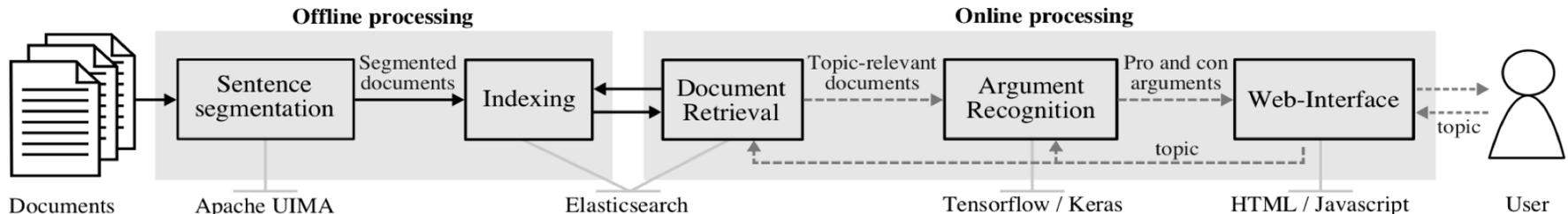
Topic relevant to the sentence



Topic **not** relevant to the sentence



Online Argument Search System



Data

- Web corpus (CommonCrawl)
- 400 Mio. English webpages

Offline Processing

- Boilerplate removal
- Sentence splitting
- Indexing using Elasticsearch

Online Processing

- Retrieve topic relevant documents
- Extract pro and con arguments

Web-Interface

- Pro/Con lists
- Source filtering
- Document ranking based on #arguments

Compare system outcome with arguments from debate portals

- Three topics from ProCon.org (“cellphones”, “social networking”, and “animal testing”)
- 1,529 classified sentence from our system
- Three undergraduate students of computer science

For each sentence s from our system

- Can s be mapped to an expert-created argument (coverage)?
- Is s a completely new argument (novelty)?
- Is s not an argument / wrong stance / nonsensical (no argument)?

Results

- Coverage: 89% with arguments from ProCon.org (full coverage for two topics)
- Novelty: 12% are completely new arguments not mentioned on ProCon.org
- No argument: 47% are either an argument classified with a wrong stance, a non-argument, or nonsensical

Summary

Sentential annotation model

- Reliably applicable to heterogeneous texts
- Simple enough for crowdsourcing

New corpus for argument search

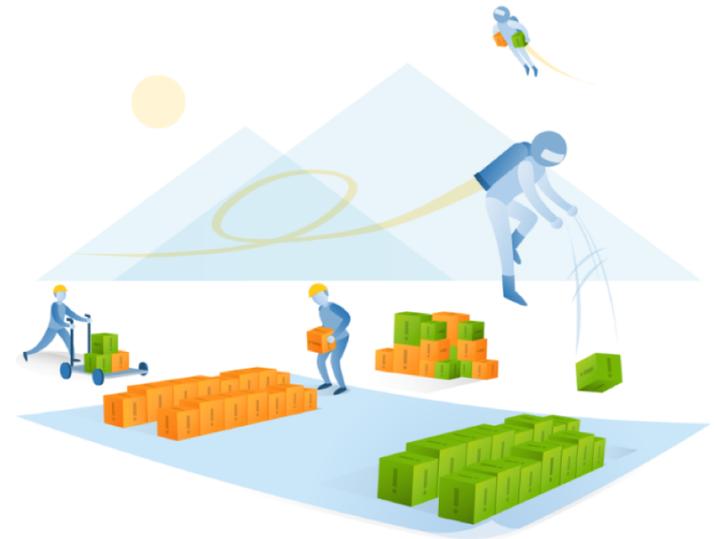
- Heterogeneous text types
- Allows cross-topic experiments

Cross-topic experiments

- Inner-att+cos generalizes best
- Achieves almost in-domain results when trained with additional topics

Future Work

- Language adaptation to support German, argument clustering and structuring



Thank you for your attention.



ArgumenText

Researchers involved in this project (alphabetical order)



Johannes
Daxenberger



Steffen
Eger



Tristan
Miller



Benjamin
Schiller



Christian
Stab



Chris
Stahlhut



Christopher
Tauchmann