Claim Detection
from middle and high school student essays

Tariq Alhindi
Debanjan Ghosh
Motivation

• The central component of the argument is the claim
  “an assertion that deserves our attention” (Toulmin, 2003)
  “a statement that is in dispute and that we are trying to support with reasons” (Govier, 2010)

• Argumentation
  segmentation, component (claim, premise) and relation (support, attack) detection
  claim example “Parents should be limiting screen time for their children”

• Argument/Claim Detection applications
  Assess public opinion on political and social issues to foster public deliberation
  Refine search and information retrieval
  other applications: legal documents, fact-checking
Motivation

Educational Applications

Analyze students' writings for scoring (Klebanov et al. 2014, Ghosh et al. 2016)

Generate quantitative and qualitative feedbacks to help students (K-12) writings such as on the Writing Mentor app
Writing Mentor App

• Essay

There must be stricter punishments to countries or people who break them. War is always brutal, but this will at least reduce the number of casualties. If we could prevent war crimes, then many more people would live through a conflict. Specific weapons being banned for reasons like massive property damage, unhuman deaths, can seriously injure someone, uncontainable, and uncontrollable. The land will suffer. A good example of the land suffer in a war is agent orange during Vietnam. Agent Orange has had adverse effects on Vietnam’s foliage and animal life. Dioxin is a highly persistent chemical compound that lasts for many years in the environment, particularly in soil, lake and river sediments and in the food chain. 60% of deaths in WW2 were civilians. Civilian deaths totaled 50 - 55 million in WW2. World War II Casualties. In Germany alone had over 8,680,000 civilian deaths. Digging trenches caused trampling of grassland, crushing of plants and animals, and churning of soil. Erosion resulted from forest logging to expand the network of trenches.

https://mentormywriting.org/
There must be stricter punishments to countries or people who break them. War is always brutal, but this will at least reduce the number of casualties. If we could prevent war crimes, then many more people would live through a conflict. Specific weapons being banned for reasons like massive property damage, inhuman deaths, can seriously injure someone, uncontainable, and uncontrollable. The land will suffer. A good example of the land suffer in a war is agent orange during Vietnam. Agent Orange has had adverse effects on Vietnam's foliage and animal life. Dioxin is a highly persistent chemical compound that lasts for many years in the environment, particularly in soil, lake and river sediments and in the food chain. 60% of deaths in WW2 were civilians. Civilian deaths totaled 50 - 55 million in WW2. World War II Casualties. In Germany alone had over 8,680,000 civilian deaths. Digging trenches caused trampling of grassland, crushing of plants and animals, and churning of soil. Erosion resulted from forest logging to expand the network of trenches.
Research questions

Can we detect claim boundaries at the token level from student essays?

If we could prevent war crimes, then many more people would live through a conflict.

War is always brutal, but this will at least reduce the number of casualties.

Can we develop robust claim detection models that go beyond lexicon-based matching?
Task

Claim Token Detection in each sentence using a B-I-O setup

- **B** beginning of a claim
- **I** inside a claim
- **O** outside a claim

If we could prevent war crimes, then many more people would live through a conflict.

War is always brutal, but this will at least reduce the number of casualties.
Task

Claim Token Detection in each sentence using a B-I-O setup

- B: beginning of a claim
- I: inside a claim
- O: outside a claim

If we could prevent war crimes, then many more people would live through a conflict.

War is always brutal, but this will at least reduce the number of casualties.
Outline of the Talk

Background and Data
Features and Models
Results
Discussions and Future Work
Background

Segmentation (argument token detection)
Argument Type Detection (claim, premise)
Argument Relation Detection (support, attack)
    using 400 argumentative essays (Stab & Gurevych, 2017, Eger et al. 2017))

Sentence Claim Detection
(-) a sentence might have multiple claims
(-) a sentence might have a claim and a premise

Cross-domain claim detection in 6 datasets (Daxenberger et al. 2017)
Using unlabeled data from the same domain (Chakrabarty et al. 2019)
Annotation Efforts

• Continuing annotation of argumentative and narrative essays with 3 annotators

• Annotation of claims, arguing expressions (``I agree'') and opinions (``music makes people feel good''), etc.

• Several rounds of calibration to finalize the guideline
  10 essays (argumentative and narrative) were triple annotated in the pilot task
  65 essays are annotated by pair(s); rest are annotated by one annotator

• Average pairwise agreement is in mid 80s (Wiebe et al., 2005)
# Dataset

## Argumentative

<table>
<thead>
<tr>
<th>Split</th>
<th>Essays</th>
<th>B-claim</th>
<th>I-claim</th>
<th>O-claim</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>train</td>
<td>45</td>
<td>951</td>
<td>10,263</td>
<td>25,332</td>
<td>36,546</td>
</tr>
<tr>
<td>test</td>
<td>45</td>
<td>707</td>
<td>7,407</td>
<td>16,841</td>
<td>24,955</td>
</tr>
</tbody>
</table>

## Narrative

<table>
<thead>
<tr>
<th>Split</th>
<th>Essays</th>
<th>B-claim</th>
<th>I-claim</th>
<th>O-claim</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>test</td>
<td>45</td>
<td>121</td>
<td>34,879</td>
<td>21,465</td>
<td>36,195</td>
</tr>
</tbody>
</table>
Features

• Discrete Features
• Word Embeddings from Transformers (e.g. BERT)

Models

• Logistic Regression (Token Classification)
• Conditional Random Field (Sequence Classification)
• Deep Learning Models (Transformers)
Discrete Feature Groups

**Structural**
- token position features, punctuation features
- position of covering sentence

**Syntactic**
- POS, Lowest common ancestor (LCA), LCA types

**Lex-Syntactic**
- Relations governing the token and its context
- N hops deep in retrieving relations
- Features: (token, previous, next): token_dependency-relation

Feature-based Models

**Logistic Regression (LR)**
Takes a single token as input.

**Conditional Random Field (CRF)**
Takes a sequence as input (sentence)
Make prediction to each element (token) in the sequence while considering other neighboring elements (tokens in the same sentence)
Deep Learning Models

• Training a Neural Network, from scratch with layers such as:
  - Dense (fully-connected)
  - Convolutional (zonal information; filters)
  - Recurrent (sequential)

• Pretrained transformers (e.g., BERT) as feature extractors
  - language models trained on a lot of data (wikipedia, google books)
  - using embeddings generated by these models as features

• Pretrained transformers as classifiers
  - fine-tuning on the task (training partition)
  - fine-tuning on a related unlabeled corpus
Fine-tuning Transformers

Training a Bert-based classifier

1. Fine-tuning on the training partition
   model choice and hyperparameter tuning (cased vs uncased, batch size, number of epochs, etc.), weighted loss per class according to class frequencies
Fine-tuning Transformers

Training a Bert-based classifier

1. Fine-tuning on the training partition:
   - model choice and hyperparameter tuning (cased vs uncased, batch size, number of epochs, etc.), weighted loss per class according to class frequencies

2. Language model fine-tuning on argumentative corpora:
   - IMHO: trained on argumentative subreddit (Chakrabarty et al. 2019)
   - GRE essays: trained on student essays (Ghosh et al. 2020)
Outline of the Talk

Background and Data

Features and Models

Results

Discussions and Future Work
Results - LR

F1 score
Results - CRF

F1 score
Results - BERT

BERT on WM argumentative

F1 score
Results - summary

F1 score
Outline of the Talk

Background and Data

Features and Models

Results

Discussions and Future Work
Model Output

Gold

There must be stricter punishments to countries or people who break them. War is always brutal, but this will at least reduce the number of casualties. If we could prevent war crimes, then many more people would live through a conflict. Specific weapons being banned for reasons like massive property damage, unhuman deaths, can seriously injure someone, uncontrollable, and uncontrollable. The land will suffer. A good example of the land suffer in a war is agent orange during Vietnam. Agent Orange has had adverse effects on Vietnam's foliage and animal life. Dioxin is a highly persistent chemical compound that lasts for many years in the environment, particularly in soil, lake and river sediments and in the food chain. 60% of deaths in WW2 were civilians. Civilian deaths totaled 50 - 55 million in WW2. World War II Casualties. In Germany alone had over 8,680,000 civilian deaths. Digging trenches caused trampling of grassland, crushing of plants and animals, and churning of soil. Erosion resulted from forest logging to expand the network of trenches.

LR (all-discrete)

There must be stricter punishments to countries or people who break them. War is always brutal, but this will at least reduce the number of casualties. If we could prevent war crimes, then many more people would live through a conflict. Specific weapons being banned for reasons like massive property damage, unhuman deaths, can seriously injure someone, uncontrollable, and uncontrollable. The land will suffer. A good example of the land suffer in a war is agent orange during Vietnam. Agent Orange has had adverse effects on Vietnam's foliage and animal life. Dioxin is a highly persistent chemical compound that lasts for many years in the environment, particularly in soil, lake and river sediments and in the food chain. 60% of deaths in WW2 were civilians. Civilian deaths totaled 50 - 55 million in WW2. World War II Casualties. In Germany alone had over 8,680,000 civilian deaths. Digging trenches caused trampling of grassland, crushing of plants and animals, and churning of soil. Erosion resulted from forest logging to expand the network of trenches.

WM App

There must be stricter punishments to countries or people who break them. War is always brutal, but this will at least reduce the number of casualties. If we could prevent war crimes, then many more people would live through a conflict. Specific weapons being banned for reasons like massive property damage, unhuman deaths, can seriously injure someone, uncontrollable, and uncontrollable. The land will suffer. A good example of the land suffer in a war is agent orange during Vietnam. Agent Orange has had adverse effects on Vietnam's foliage and animal life. Dioxin is a highly persistent chemical compound that lasts for many years in the environment, particularly in soil, lake and river sediments and in the food chain. 60% of deaths in WW2 were civilians. Civilian deaths totaled 50 - 55 million in WW2. World War II Casualties. In Germany alone had over 8,680,000 civilian deaths. Digging trenches caused trampling of grassland, crushing of plants and animals, and churning of soil. Erosion resulted from forest logging to expand the network of trenches.

CRF (all-discrete + embeddings)

There must be stricter punishments to countries or people who break them. War is always brutal, but this will at least reduce the number of casualties. If we could prevent war crimes, then many more people would live through a conflict. Specific weapons being banned for reasons like massive property damage, unhuman deaths, can seriously injure someone, uncontrollable, and uncontrollable. The land will suffer. A good example of the land suffer in a war is agent orange during Vietnam. Agent Orange has had adverse effects on Vietnam's foliage and animal life. Dioxin is a highly persistent chemical compound that lasts for many years in the environment, particularly in soil, lake and river sediments and in the food chain. 60% of deaths in WW2 were civilians. Civilian deaths totaled 50 - 55 million in WW2. World War II Casualties. In Germany alone had over 8,680,000 civilian deaths. Digging trenches caused trampling of grassland, crushing of plants and animals, and churning of soil. Erosion resulted from forest logging to expand the network of trenches.
Model Output

Gold

There must be stricter punishments to countries or people who break them. War is always brutal, but this will at least reduce the number of casualties. If we could prevent war crimes, then many more people would live through a conflict. Specific weapons being banned for reasons like massive property damage, unhuman deaths, can seriously injure someone, uncontainable, and uncontrollable. The land will suffer. A good example of the land suffer in a war is agent orange during Vietnam. Agent Orange has had adverse effects on Vietnam's foliage and animal life. Dioxin is a highly persistent chemical compound that lasts for many years in the environment, particularly in soil, lake and river sediments and in the food chain. 60% of deaths in WW2 were civilians. Civilian deaths totaled 50 - 55 million in WW2. World War II Casualties. In Germany alone had over 8,680,000 civilian deaths. Digging trenches caused trampling of grassland, crushing of plants and animals, and churning of soil. Erosion resulted from forest logging to expand the network of trenches.

WM App

There must be stricter punishments to countries or people who break them. War is always brutal, but this will at least reduce the number of casualties. If we could prevent war crimes, then many more people would live through a conflict. Specific weapons being banned for reasons like massive property damage, unhuman deaths, can seriously injure someone, uncontainable, and uncontrollable. The land will suffer. A good example of the land suffer in a war is agent orange during Vietnam. Agent Orange has had adverse effects on Vietnam's foliage and animal life. Dioxin is a highly persistent chemical compound that lasts for many years in the environment, particularly in soil, lake and river sediments and in the food chain. 60% of deaths in WW2 were civilians. Civilian deaths totaled 50 - 55 million in WW2. World War II Casualties. In Germany alone had over 8,680,000 civilian deaths. Digging trenches caused trampling of grassland, crushing of plants and animals, and churning of soil. Erosion resulted from forest logging to expand the network of trenches.

CRF (all-discrete + embeddings)

There must be stricter punishments to countries or people who break them. War is always brutal, but this will at least reduce the number of casualties. If we could prevent war crimes, then many more people would live through a conflict. Specific weapons being banned for reasons like massive property damage, unhuman deaths, can seriously injure someone, uncontainable, and uncontrollable. The land will suffer. A good example of the land suffer in a war is agent orange during Vietnam. Agent Orange has had adverse effects on Vietnam's foliage and animal life. Dioxin is a highly persistent chemical compound that lasts for many years in the environment, particularly in soil, lake and river sediments and in the food chain. 60% of deaths in WW2 were civilians. Civilian deaths totaled 50 - 55 million in WW2. World War II Casualties. In Germany alone had over 8,680,000 civilian deaths. Digging trenches caused trampling of grassland, crushing of plants and animals, and churning of soil. Erosion resulted from forest logging to expand the network of trenches.

BERT (fine-tuned with GRE data)

There must be stricter punishments to countries or people who break them. War is always brutal, but this will at least reduce the number of casualties. If we could prevent war crimes, then many more people would live through a conflict. Specific weapons being banned for reasons like massive property damage, unhuman deaths, can seriously injure someone, uncontainable, and uncontrollable. The land will suffer. A good example of the land suffer in a war is agent orange during Vietnam. Agent Orange has had adverse effects on Vietnam's foliage and animal life. Dioxin is a highly persistent chemical compound that lasts for many years in the environment, particularly in soil, lake and river sediments and in the food chain. 60% of deaths in WW2 were civilians. Civilian deaths totaled 50 - 55 million in WW2. World War II Casualties. In Germany alone had over 8,680,000 civilian deaths. Digging trenches caused trampling of grassland, crushing of plants and animals, and churning of soil. Erosion resulted from forest logging to expand the network of trenches.
Observations

• B-claim after back B-claim
• I-claim starts after O-claim

post-processing based on the possible transitions

B → I
I → B  I → I  I → O
O → B  O → O

• Multiple claims in a sentences are sometimes merged
  comma are sometimes within a claim or separate two claims
Challenges

• Claims often look like other type of argumentative text expressions ("I agree") premise or opinions

• Run-on sentences and other writing features specific to middle or high school students

• Genre/domain differences (argumentative vs narrative) makes generalizability and data augmentation harder
Future Work

• Using RNN (+CRF) with word embeddings and discrete features

• Joint Learning
  Argument token detection (claim+premise, claim only)
  Claim sentence detection
  Multi-dataset training

• Pretraining on a relevant task
  SG2017 vs WM2020
  Argumentative vs Narrative
Educational Applications

• Correlation claims are with scores

• Useful for feedback generation

Great job!
Your essay have 8 supported claims
Thank You