Towards Effective Rebuttal:
Listening Comprehension using Corpus-Wide Claim Mining

Tamar Lavee[*][†]**, Matan Orbach**[*]**, Lili Kotlerman, Yoav Kantor, Shai Gretz, Lena Dankin,**
Shachar Mirkin, Michal Jacovi, Yonatan Bilu, Ranit Aharonov and Noam Slonim
IBM Research

Abstract

Engaging in a live debate requires, among other things, the ability to effectively rebut arguments claimed by your opponent. In particular, this requires identifying these arguments. Here, we suggest doing so by automatically mining claims from a corpus of news articles containing billions of sentences, and searching for them in a given speech. This raises the question of whether such claims indeed correspond to those made in spoken speeches. To this end, we collected a large dataset of 400 speeches in English discussing 200 controversial topics, mined claims for each topic, and asked annotators to identify the mined claims mentioned in each speech. Results show that in the vast majority of speeches debaters indeed make use of such claims. In addition, we present several baselines for the automatic detection of mined claims in speeches, forming the basis for future work. All collected data is freely available for research.

1 Introduction

Project Debater[1] is a system designed to engage in a full live debate with expert human debaters. One of the major challenges in such a debate is listening to a several-minute long speech delivered by your opponent, identifying the main arguments, and rebutting them with effective persuasive counter arguments. This work focuses on the former, namely, automatically identifying arguments mentioned in opponent speeches.

One of the fundamental capabilities developed in Debater is the automatic mining of claims (Levy et al., 2014) – general, concise statements that directly support or contest a given topic – from a large text corpus. It allows Debater to present high-quality content supporting its side within its generated speeches. Our approach utilizes this capability for a different purpose: claims mined *from the opposing side* are searched for in a given opponent speech.

The implicit assumption in this approach is that mined claims would be often said by human opponents. This is far from trivial, since mined content from a large text corpus is not guaranteed to provide enough coverage over arguments made by individual human debaters. To assess this, we collected a large and varied dataset of recorded speeches discussing controversial topics, along with an annotation specifying which mined claims are mentioned in each speech.

Annotation results show our approach obtains good coverage, thus making the task of *claim matching* – automatically identifying given claims in speeches – interesting in the context of *mined* claims. Using the collected data, several claim matching baselines are examined, forming the basis for future work in this direction.

The main contributions of this paper are: (i) a recorded dataset of 400 speeches discussing 200 controversial topics, along with mined claims for each topic; (ii) an annotation specifying the claims mentioned in each speech; (iii) baselines for matching mined claims to speeches. All collected data is freely available for further research[2].

2 Related Work

Mirkin et al. (2018b) recently presented a dataset similar to the one we collected in the context of Machine Listening Comprehension (MLC) over argumentative content. Instead of using mined claims, they extracted lists of potential arguments

---

[1]www.research.ibm.com/
artificial-intelligence/project-debater

[2]https://www.research.ibm.com/haifa/
dept/vst/debating_data.shtml

from iDebate[3], a manually curated high-quality database containing arguments for controversial topics. A major drawback of such an approach is topic coverage – any topic not included in the database cannot be handled. Another limitation is that argument lists from iDebate are short, each typically contains only 3 or 4 arguments from each side.

MLC has been recently gaining attention, and there are several new interesting works and datasets (Lee et al., 2018b,a; Ünlü et al., 2019). Other tasks are often phrased as a collection of test questions, which can be multiple choice (Tseng et al., 2016; Fang et al., 2016) or require, for example, identifying an entity mentioned by the speaker (Surdeanu et al., 2006; Comas et al., 2010).

Methods for detecting claims in given texts have been applied to various argumentative domains (e.g. by Palau and Moens (2011); Stab and Gurevych (2017); Habernal and Gurevych (2017)). While such tools may be applied to opponent speeches, a major difference in our setting is that it involves *spoken* rather than *written* language. Spoken spontaneous speeches often contain disfluencies such as breaks, repetitions, or other irregularities, and therefore claims detected in spoken content are likely to contain them as well. In addition, since the opponent speech audio is transcribed into text using an Automatic Speech Recognition (ASR) system, its errors propagate to detected claims. This is a crucial point for Debater – since a desired rebuttal in live debates typically includes a quote of the argument made by the opponent. Thus, any single disfluency or ASR error in a detected claim prevents its actual use.

3 Data

Motions As in Mirkin et al. (2018b), we manually curated a list of 200 controversial topics - referred to as motions, as in formal parliamentary proposals. Each motion focuses on a single Wikipedia concept, and is phrased similarly to parliamentary motions, e.g. *We should introduce compulsory voting*.

Speeches For each motion we recorded two argumentative speeches *contesting* it, as described in Mirkin et al. (2018b), producing a total of 400 speeches. Our choice of recording speeches contesting (rather than supporting) the motion is arbi-

trary, and all methods described henceforth would work similarly on speeches recorded for the other side. The dataset format follows the one described in Mirkin et al. (2018a). Each speech is associated with a corresponding audio file, an automatic transcription of it[4], and a manually-transcribed "reference" text. Speeches were recorded by 9 expert debaters. On average, a speech contains 29 sentences and 748 tokens. The average ASR word error rate, computed by comparing to the manual transcripts, is 7.07%.

Mining Claims Figure 1 illustrates the suggested mined–claims based rebuttal generation pipeline. Following is a brief description of the existing components which perform claim mining. The rest of this work focuses on the subsequent component which identifies mentioned claims in speeches.

Processing starts from a large corpus of news articles containing billions of sentences. Given a controversial topic, several queries are applied, retrieving sentences which potentially contain claims that are relevant to the topic. Query results are then ranked by a neural-model trained to detect sentences containing claims (similarly to Levy et al. (2017, 2018)[5]). Top-ranked sentences are passed to a boundary detection component, responsible for finding the exact span of each claim within each sentence (Levy et al., 2014). Lastly, the stance of each claim towards the topic is detected using the method of Bar-Haim et al. (2017). Used models are tuned towards precision, aimed at obtaining a set of coherent, grammatically–correct claims from the opponent side, which can then be directly quoted in a live debate.

Prior to claim matching, mined claims are filtered, aiming to focus on those with a higher chance of obtaining a successful match. This included removing claims containing: (i) more than 10 tokens, since longer claims are less concise and may contain more than a single idea; (ii) named entities (found with Stanford NER (Finkel et al., 2005)), other than the topic itself, assuming they are too specific; (iii) unresolved demonstratives, which may hint to an incoherent sentence or an error in boundary detection.

The released dataset includes all output from these components, as well as a complete labeling
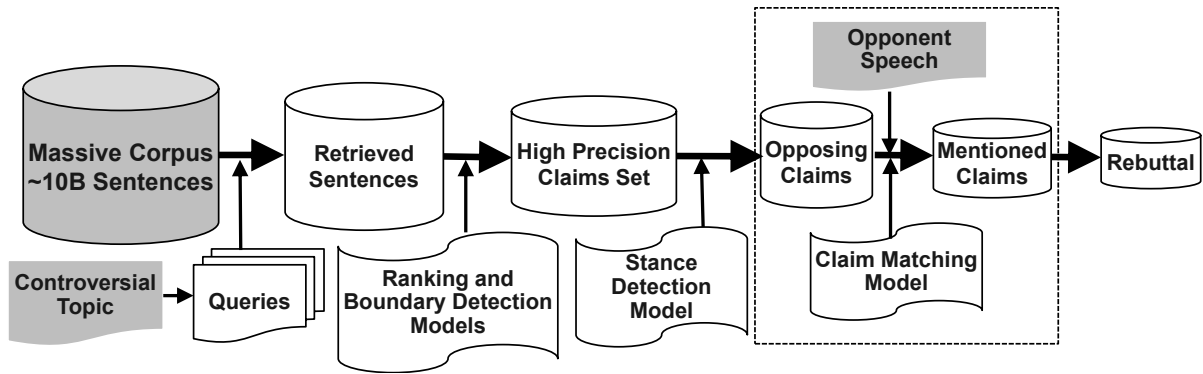

Figure 1: The suggested architecture for mined–claims based rebuttal generation. System inputs are depicted with a gray background. The focus of this work is marked on the right: detecting mentioned claims in opponents speeches. Preceding existing components are shortly described in Section 3. The entire pipeline starts from billions of sentences, and its final goal is producing few high quality rebuttals opposing the opponent speech.

indicating which texts are erroneously predicted to be claims, and what is the correct stance of all valid claims. The percentage of mined texts which are both labeled as claims and have a correctly identified stance is 86%.

Topic coverage   Claim mining yielded, on average, 12.2 claims for each speech, suggesting match-candidates for 93.5% of the motions in our data. This shows the potentially high coverage of using mined claims. In contrast, only 39% of these motions have candidate iDebate arguments present in the dataset of Mirkin et al. (2018b).

4   Annotation

Next, we assessed whether mined claims are mentioned in recorded speeches through annotation. In case mined claims do occur in many speeches, the collected labels would form a dataset which can be used to develop algorithms for identifying mined claims in speeches.

In our annotation scheme, each question included a speech followed by a list of mined claims, and we asked to mark those claims which were mentioned by the speaker. Speeches were given in both text (manual transcription) and audio formats, to allow for listening, reading, or both. The length of each claim list was limited to at most 20 claims. Longer lists were split into multiple questions for the same speech.

Initially the task allowed for two labels: *Mentioned* or *Not mentioned*, yet error analysis showed major disagreements on claims alluded to, but not explicitly stated, in a speech. Example 1 illustrates this for the claim *compulsory voting is undemocratic*. Some annotators considered such cases

as mentioned, while others disagreed. Thus, we modified the task to include three labels (*Explicit*, *Implicit*, *Not mentioned*), and provided detailed examples in the guidelines. Example 1 further shows an explicit mention of the same claim[6].

Example 1 (Implicit / explicit mentions)
Claim: *Compulsory voting is undemocratic*

Implicit ...*people have a right to not vote ... that's the way that rights work ... if you think that there is literally any reason a person might not want to vote ... you should ensure that that person is not penalized for not voting...*

Explicit ...*it might be preferable if everyone voted, but it is undemocratic to force everyone to vote.*

Quality control   Annotation of each question is time-consuming, since it requires going over a whole speech, and a list of claims. Combined with the amount of questions, we resorted to working with a crowd-sourcing platform[7], to make annotation practical. This required close monitoring and the removal of unreliable annotators. For quality control, we placed "test" claims among real mined claims, either using claims from different motions, expecting a negative answer, or by using claims unanimously labeled as mentioned for the same speech in previous rounds, expecting a positive label (explicit or implicit). We then defined thresholds on the accuracy of labeling of these test claims, and on the agreement of an annotator with its peers, disqualifying those who did not meet them. In addition, good annotators were awarded

---

[6]Full annotation guidelines, including more examples, are provided in the Appendix.
[7]Figure-Eight: www.figure-eight.com

bonus payments, in order to keep them engaged. Each question was answered by seven annotators.

Annotation results A claim is considered as *mentioned* in a speech when a majority of annotators marked it as either an explicit or an implicit mention. A mentioned claim is an *explicit* mention when its explicit answer count is strictly larger than its implicit answer count. Otherwise, it is an *implicit* mention.

Overall, annotation of all 400 speeches and their mined claims amounted to 4,882 speech–claim pairs. Of these, 34.7% were annotated as claims mentioned in the speech. Only 5.6% are explicit mentions, testifying to the difficulty of the matching task.

On average, there were 4.2 mentioned claims in every speech. 82.5% of the labels were agreed on by at least 5 out of the 7 annotators. The percentage of claims mentioned at least once is 44.8%, and in 87.3% of speeches at least one claim is mentioned (6.5% of speeches had no mined claims).

Annotation Quality To estimate inter-annotator agreement, we focus on annotators with a significant contribution, selecting those who have answered more than 20 common questions with each of at least 5 different peers. A per-annotator agreement score is defined by averaging Cohen's Kappa (Cohen, 1960) calculated with each peer. The final agreement score is the average of all annotators agreement scores.

Considering two labels (mentioned or not), agreement was 0.44. Mirkin et al. (2018b) reported a score of 0.5 on a similar annotation scheme performed by expert annotators. The difference is potentially due to the use of crowd, and the larger group of annotators taking part.

Note the applicability of chance-adjusted agreement scores to the crowd has been questioned, in particular for tasks within the argumentation domain (Passonneau and Carpenter, 2014; Habernal and Gurevych, 2016). Our test claims allow further validation of annotation quality, since their answers are known a-priory. The average annotator error rate on those test claims is low: 7.8%.

5 Evaluation

Annotation confirmed our hypothesis that claims mined from a corpus are indeed mentioned, or are at least alluded to, in spontaneous speeches on controversial topics. On average, of the 12.2 claims mined for each speech, about a third were annotated as mentioned. We now present several baselines for identifying those mentioned claims, using the collected data.

Speech pre–processing An input audio speech is automatically transcribed into text using IBM Watson ASR[8]. The text is then segmented to sentences as in Pahuja et al. (2017).

Next, given a claim, semantically similar sentences are identified. Each sentence is represented using a 200-dimensional vector constructed by: removing stopwords; representing remaining words using word2vec (*w2v*) (Mikolov et al., 2013) word embeddings learned over Wikipedia; computing a weighted average of those word embeddings using tf-idf weights (idf values are counted when considering each Wikipedia sentence as a document). The claim is represented similarly, and its semantic similarity to a given sentence is computed using the cosine similarity between their vector representations. All sentences with low similarity to the claim are ignored (using a fixed threshold).

Remaining sentences are scored by the harmonic mean (*HM*) of three additional semantic similarity measures, and the top-K ranked sentence are selected (we experiment with K ∈ $\{1, 3, 5\}$). These features are:
– *Concept Coverage*: The fraction of Wikipedia concepts identified in the claim, found within the sentence.
– *Parse Pairs*: The parse trees of the claim and the sentence are obtained using Stanford parser (Socher et al., 2013). Then, pairwise edge similarity is defined to be the harmonic mean of the cosine similarities computed between the two parent word embeddings and the two child word embeddings. Each edge in the claim parse tree is scored using its maximal similarity to an edge from the sentence parse tree. Averaging these scores yields the final feature score.
– *Explicit Semantic Analysis* (Gabrilovich and Markovitch, 2007): Cosine similarity computed between vector representations of the claim and sentence over the Wikipedia concepts space.

Methods Following sentence selection, three methods are considered for scoring a speech and

---

[8]www.ibm.com/watson/services/speech-to-text

a claim:

HM: Averaging the selected sentences *HM* scores.
NN: Using a Siamese Network (Bromley et al., 1993), containing K instances of the same sub-network: Each pair of a selected sentence and the claim is embedded with a BiLSTM, followed by an attention layer, a fully connected layer, and finally a *softmax* layer which yields a score for the pair. The network outputs the maximum score of these K sub-networks.
LR: calculating 23 similarity measures between each selected sentence and the claim. For each measure, the average over the K selected sentences is taken. These averages are used as features for training a logistic regression classifier. Following is a brief description of the different groups of similarity measures we used.
– **w2v-based similarities** (5 features): Computing pairwise word similarities using the cosine similarity of the corresponding word embeddings, and applying several aggregation options.
– **Parse tree similarities** (6 features): Computing the parse tree of the claim and the sentence, and calculating similarities between different elements of those trees, similarly to the *Parse Pairs* feature described above.
– **Part of speech (POS) similarities** (5 features): Identifying tokens with a specific POS tag in the texts, and computing either the fraction of such tokens from one text which appear in the other, or otherwise aggregating w2v-based cosine similarities between these tokens in several ways.
– **Wikipedia concepts similarities** (2 features): The fraction of Wikipedia concepts from the claim which are present in the sentence, and vice versa.
– **Lexical similarities** (5 features): n-grams are extracted from the two texts in various settings (e.g. with or without lemmatization, or using different values of n). Then, each n-gram from the claim is scored by its maximal similarity to sentence n-grams (using a w2v-based similarity, with tf/idf weights). The feature values is the average of these scores.

Training and test sets The data was randomly split into a *train* and *test* sets, equal in size. Each contains 100 motions and 200 speeches. The number of labeled speech-claim pairs is 2,456 in *train* and 2,426 in *test*.

Model selection as well as hyper-parameters tuning, such as the selection of K, are performed on *train* (using cross validation for *LR* and *NN*).
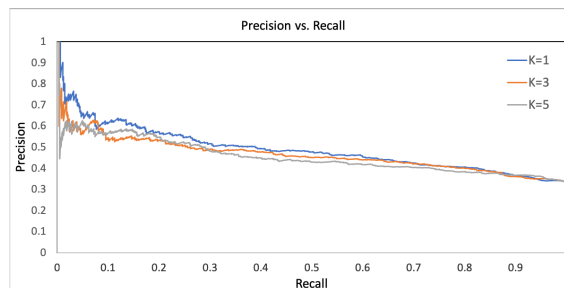

Figure 2: Precision-Recall curves for the top-3 claim matching configurations (all using *HM*) on *test*.

Different configuration are ranked according to their Area Under the ROC Curve (AUC) measure.

Results The AUC score of both *LR* and *NN* on *train*, for various values of K, was no higher than 0.57. In contrast, all *HM* configurations achieved AUC higher than 0.62. We therefore focus on this method, though it is interesting, in future work, to improve the supervised methods or understand why they work somewhat poorly. Figure 2 shows precision-recall curves for *HM* and the different values of K on *test*. The different plots are comparable, yet there is a slight advantage to K = 1 for applications valuing precision over recall.

6 Conclusions and Future Work

We addressed the task of identifying arguments claimed in spoken argumentative content. Our suggested approach utilized claims mined from a large text corpora.The collected labeled data show these claims do cover, in most cases, arguments made by expert debaters. This confirms this is a valid approach for solving this task.

Interestingly, most claims are made implicitly, suggesting that assertion of claims often involves high lexical variability and expression of ideas across multiple (not always consecutive) sentences. This poses a challenge for automatic claim matching methods, as made evident by the baselines discussed here.

Successfully identifying arguments made by opponents forms the basis for an effective rebuttal. Our work leaves open the question of how to construct such rebuttals once a claim has been matched. This would be an interesting research direction for future work.

7 Acknowledgments

References

Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. 2017. 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, Valencia, Spain. Association for Computational Linguistics.

Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. 1993. Signature verification using a "siamese" time delay neural network. In *Proceedings of the 6th International Conference on Neural Information Processing Systems*, NIPS'93, pages 737–744, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.

Pere Comas, Jordi Turmo, and Lluís Màrquez. 2010. Using dependency parsing and machine learning for factoid question answering on spoken documents. In *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010*, pages 1265–1268.

Wei Fang, Juei-Yang Hsu, Hung-yi Lee, and Lin-Shan Lee. 2016. Hierarchical attention model for improved machine comprehension of spoken content. In *2016 IEEE Spoken Language Technology Workshop, SLT 2016, San Diego, CA, USA, December 13-16, 2016*, pages 232–238.

Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 363–370, Stroudsburg, PA, USA. Association for Computational Linguistics.

Evgeniy Gabrilovich and Shaul Markovitch. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, pages 1606–1611.

Ivan Habernal and Iryna Gurevych. 2016. Which argument is more convincing? analyzing and predicting convincingness of web arguments using bidirectional lstm. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1589–1599.

Ivan Habernal and Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. *Computational Linguistics*, 43(1):125–179.

Chia-Hsuan Lee, Shang-Ming Wang, Huan-Cheng Chang, and Hung-Yi Lee. 2018a. ODSQA: Open-Domain Spoken Question Answering Dataset. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 949–956. IEEE.

Chia-Hsuan Lee, Szu-Lin Wu, Chi-Liang Liu, and Hung-yi Lee. 2018b. Spoken SQuAD: A Study of Mitigating the Impact of Speech Recognition Errors on Listening Comprehension. In *Proceedings of Interspeech*.

Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1489–1500. Dublin City University and Association for Computational Linguistics.

Ran Levy, Ben Bogin, Shai Gretz, Ranit Aharonov, and Noam Slonim. 2018. Towards an argumentative content search engine using weak supervision. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2066–2081. Association for Computational Linguistics.

Ran Levy, Shai Gretz, Benjamin Sznajder, Shay Hummel, Ranit Aharonov, and Noam Slonim. 2017. Unsupervised corpus-wide claim detection. In *Proceedings of the 4th Workshop on Argument Mining, ArgMining@EMNLP 2017, Copenhagen, Denmark, September 8, 2017*, pages 79–84.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

Shachar Mirkin, Michal Jacovi, Tamar Lavee, Hong-Kwang Kuo, Samuel Thomas, Leslie Sager, Lili Kotlerman, Elad Venezian, and Noam Slonim. 2018a. A recorded debating dataset. In *Proceedings of LREC*.

Shachar Mirkin, Guy Moshkowich, Matan Orbach, Lili Kotlerman, Yoav Kantor, Tamar Lavee, Michal Jacovi, Yonatan Bilu, Ranit Aharonov, and Noam Slonim. 2018b. Listening comprehension over argumentative content. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 719–724. Association for Computational Linguistics.

Vardaan Pahuja, Anirban Laha, Shachar Mirkin, Vikas Raykar, Lili Kotlerman, and Guy Lev. 2017. Joint Learning of Correlated Sequence Labelling Tasks Using Bidirectional Recurrent Neural Networks. *Proceedings of Interspeech*.

Raquel Mochales Palau and Marie-Francine Moens. 2011. Argumentation mining. *Artif. Intell. Law*, 19(1):1–22.

Rebecca J Passonneau and Bob Carpenter. 2014. The benefits of a model of annotation. *Transactions of the Association for Computational Linguistics*, 2:311–326.

Richard Socher, John Bauer, Christopher D Manning, et al. 2013. Parsing with compositional vector grammars. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 455–465.

Christian Stab and Iryna Gurevych. 2017. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, 43(3):619–659.

Mihai Surdeanu, David Dominguez-Sal, and Pere Comas. 2006. Design and performance analysis of a factoid question answering system for spontaneous speech transcriptions. In *INTERSPEECH 2006 - ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA, September 17-21, 2006*.

Bo-Hsiang Tseng, Sheng-Syun Shen, Hung-Yi Lee, and Lin-Shan Lee. 2016. Towards Machine Comprehension of Spoken Content: Initial TOEFL Listening Comprehension Test by Machine. In *Proceedings of Interspeech*.

Merve Ünlü, Ebru Arisoy, and Murat Saraçlar. 2019. Question answering for spoken lecture processing. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7365–7369. IEEE.

A Annotation Guidelines

Following are the guidelines used in the annotation of mined claims to recorded speeches.

Overview

In the following task you are given a speech that contests a controversial topic. You are asked to listen to the speech and/or read the transcription, then decide whether a list of potentially related claims were mentioned by the speaker explicitly, implicitly, or not at all.

Steps

- **Listen** to the speech and/or read the transcription of the speech. Note: some speeches are transcribed automatically and may contain errors.

- **Review** the list of possibly relevant claims. Note: few of the claims might not be full sentences. Please do your best to "complete" them to claims in a common-sense manner. If the claim doesn't make any sense, select "Not mentioned".

- **Decide** based on the speech only whether the speaker agrees with each claim, and choose the appropriate answer:

    - Agree - Explicitly
    - Agree - Implicitly
    - Not Mentioned

Rules & Tips

You should ask yourself whether the statement "*The speaker argued that <claim>*" is valid or not. Note, this statement can be valid even if the speaker was stating the claim using a somewhat different phrasing in her/his speech.

Examples

Agree - Explicitly

The claim was mentioned by the speaker, but perhaps phrased differently.

- If the speaker said: *organic food is simply healthier* then she explicitly agrees with the claim **organic food products are better in health**.

- If in a speech about the topic "We should ban boxing" the speaker said: *we think regulation is simply better in this instance than a ban* then she explicitly agrees with the claim **We should not ban boxing altogether, just regulate it**.

Agree - Implicitly

The claim was not mentioned by the speaker but it is clearly implied from the speech, and we know for sure that the speaker agrees with the claim.

The claim will usually be implied in one of the following ways:

- The claim is a generalization of a claim mentioned by the speaker.

    If the speaker said: *we allow people to make these decisions even if they might be physically bad for them* then she implicitly agrees with the claim **People should have the right to choose what to do with their bodies**.

- The claim summarizes an argument made by the speaker.

    If the speaker said: *It's essential that something is done to ensure that people don't have dental problems later in life. Water fluoridation is so cheap it's almost free. There are no proven side effects, the FDA and comparable groups in Europe have done lots and lots of tests and found that water fluoridation is actually a net health good, that there's no real risk to it* then she implicitly agrees with the claim **water fluoridation is safe and effective**.

- The claim can be deduced from an argument made by the speaker.

    If the speaker said *without the needle exchange program people are still going to do heroin or other kinds of drugs anyway with dirty or less safe needles. This does lead to things like HIV getting transmitted, it leads to other diseases as well, being more likely to get transmitted* then she implicitly agrees that **needle exchange programs could reduce the spread of disease**.

The text itself must contain some indication of the implied claim. Don't choose this option if you need to make an extra logical step to conclude that the speaker agrees with the claim. For example, if the speaker said *International aid has problems, but is still valuable*, then you should not conclude

that she agrees with the claim **We should fix international aid, and not get rid of it** since she did not argue that the problems should be fixed.

Not Mentioned

The claim is not part of the speech.

For example, if the speaker said *and, yes, feminism has its flaws in the status quo ... but it can be reformed, and the tenets of equality that feminism stands for ... those tenets certainly should not be abandoned, and feminism has done a fantastic job, both historically and in the modern day, of championing those tenets.* then it can not be inferred that she agrees with the claim **We should try to fix the issues with feminism because people support it**. Although she suggests to fix the issues with feminism, she does not claim that people support it.

IMPORTANT NOTE: Your answers will be reviewed after the job is complete. We trust you to perform the task thoroughly, while carefully following the guidelines. Once your answers are determined as acceptable per our review, you might receive a bonus. Note that the bonus is given to contributors who complete at least 5 pages per job, and a higher bonus may be given to contributors who complete at least 50 pages.