Contract NLI

Guided by: Dr. Manish Shrivastava
TA: Patanjali

Team 19 (Badhum)
Viraj Shah (2023201011)
Sushrut Naik (2023201064)
Ronak Patel (2023201074)

Problem Statement

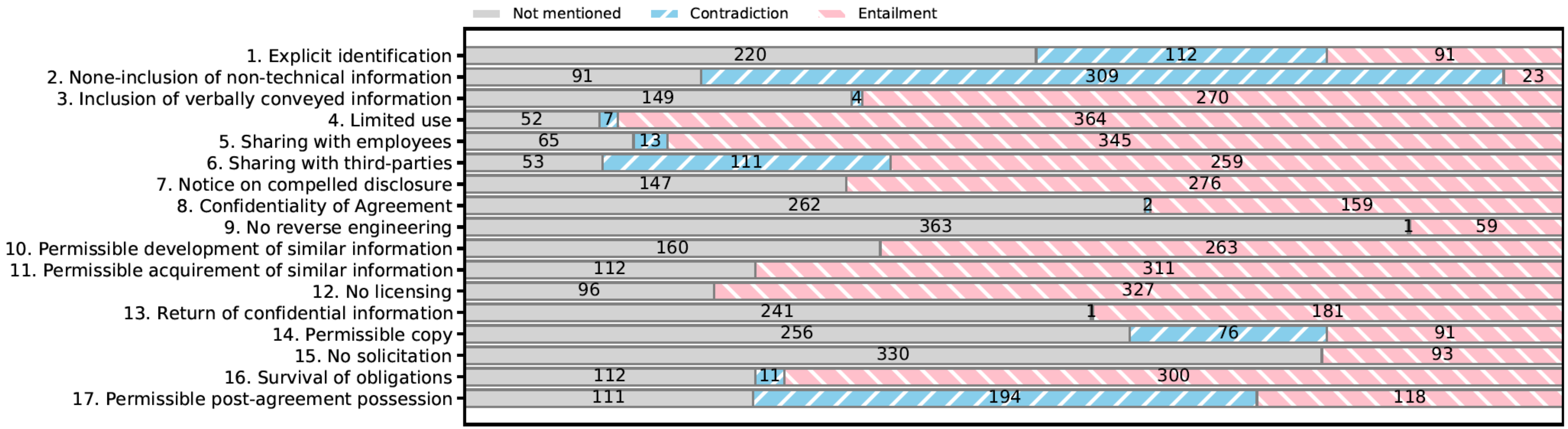
- Natural Language Inference(NLI) and evidence identification in large documents, specifically NDAs
- NLI – Given hypothesis and premise, identify relationship
    - Entailment
    - Contradiction
    - Not Mentioned
- Evidence Identification – If entailment or contradiction, which part of the premise supports that relationship

Dataset

- Statistics
  - Avg Paragraphs: 43.7
  - Avg Tokens: 2254.3
  - Avg Spans: 77.8

Format	Source	Train	Dev	Test	Total
Plain Text	EDGAR	83	12	24	119
HTML	EDGAR	79	11	23	113
PDF	Search Engines	261	38	76	375
Total		423	61	123	607

Label Distribution



Baselines

- NLI Only
  - **Majority Vote**: Assigns the majority label from the training set to each hypothesis
  - **Doc TF-IDF + SVM:** A document-level multi-class linear SVM classifier

- Evidence Identification Only
  - **Span TF-IDF + Cosine:** Identifies evidence spans using cosine similarity between hypothesis and spans
  - **Span TF-IDF + SVM:** A span-level binary SVM classifier for evidence identification

Baseline Results

	Evidence		NLI
	mAP	P@R80	Accuracy
Majority Vote	-	-	0.66
Doc TF-IDF + SVM	-	-	0.68
Span TF-IDF + Cosine	0.04	0.03	-
Span TF-IDF + SVM	0.025	0.025	-

Challenges

- Predicting start and end tokens makes the task harder by combining span boundary detection and evidence identification into a single step
- Cannot feed whole document into model (BERT: 512 tokens)
- Static windows with strides can cause spans to split across contexts or lose crucial surrounding context
- Inadequate surrounding context in split windows hinders the model’s ability to fully capture span semantics

Approach

- Motivation: Evidence spans often defined naturally, identifying boundaries not always necessary
- Model evidence identification as binary classification (is span evidence of hypothesis or not)
- Introduced dynamic context segmentation to inculcate sophisticated context

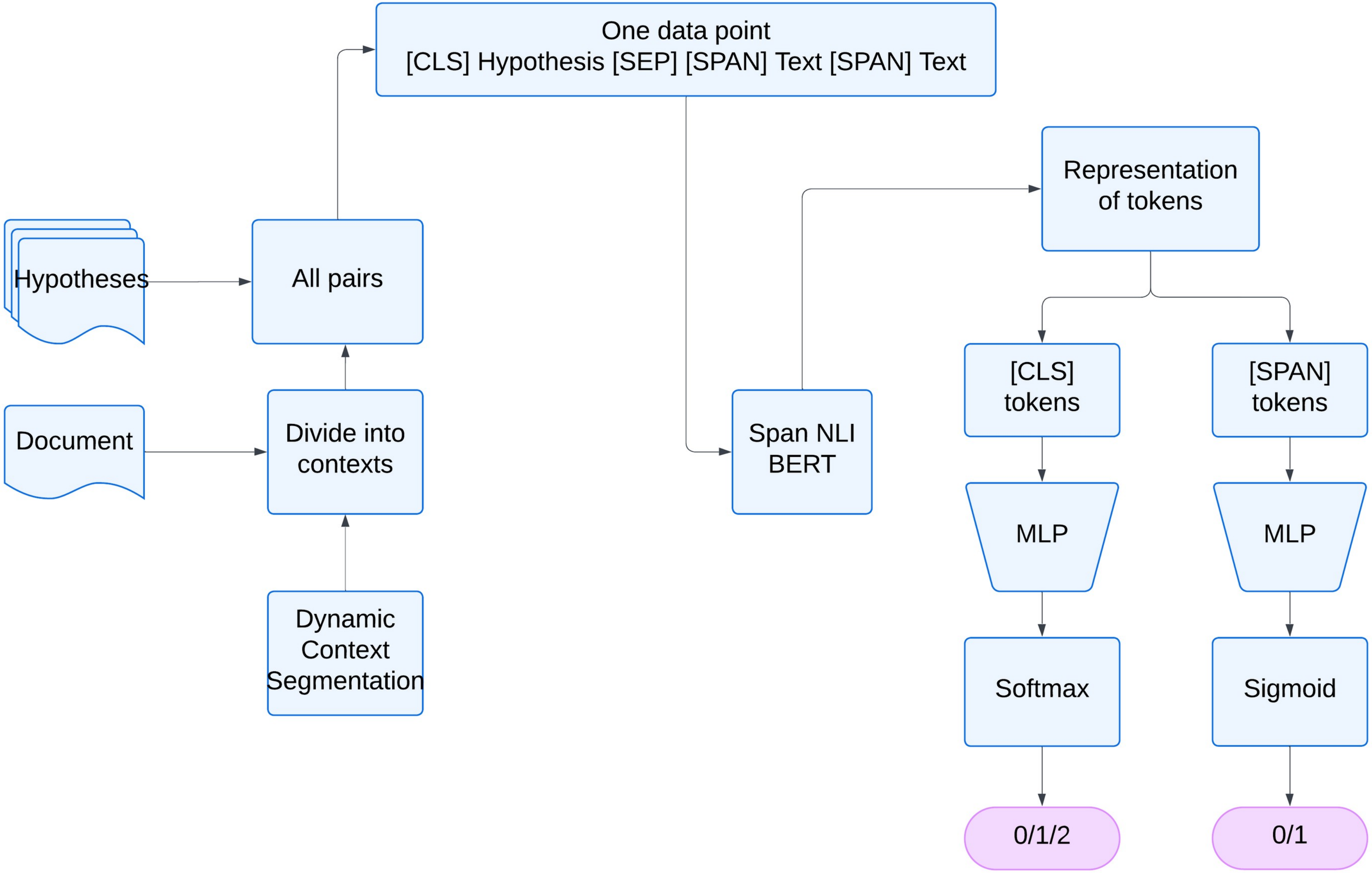
Dynamic Context Segmentation

Input: Span boundary token indices $B = [b_0, b_1, ...]$, Tokens $T = [t_0, t_1, ...]$, min. # of surrounding tokens n, max. context length l

Output: List of overlapping contexts

```
contexts = [] ;
start = 0 ;
while len(B) > 0 do
    for b_i in B where b_i − start <= l do
        B.remove(b_{i−1}) ;
        end = b_{i−1} ;
    end
    contexts.append(T[start : (start + l)]) ;
    start = end − n ;
end
return contexts ;
```

Span NLI BERT


One data point
[CLS] Hypothesis [SEP] [SPAN] Text [SPAN] Text
Representation of tokens
Hypotheses
All pairs
[CLS] tokens
[SPAN] tokens
Document
Divide into contexts
Span NLI BERT
MLP
MLP
Dynamic Context Segmentation
Softmax
Sigmoid
0/1/2
0/1

Loss Calculation

- Span Identification Loss
- NLI Loss
- Multitask Loss

Results

	Bert-base	Bert-base FFT	Bert-large	DistilBERT-FFT	DistilBERT-frozen	DistilBERT(lambda-0.05)	DistilBERT(lambda-0.2)	DistilBERT(lambda-0.4)
mAP	0.5432	0.8378	0.5883	0.8501	0.5594	0.5704	0.5708	0.5699
P@R80	0.1015	0.6778	0.2567	0.7463	0.1113	0.1205	0.1208	0.1208
ACC(NLI)	0.6054	0.6693	0.6554	0.6862	0.6059	0.6111	0.6152	0.6155
F1 (C)	0.3024	0.2550	1.0	0.2768	0.2596	0.2659	0.2899	0.2900
F1 (E)	0.2704	0.2646	0.2621	0.2948	0.2999	0.2932	0.2664	0.2656

Limitations and future work

- Imbalanced label distribution and scarcity of data
- Linguistic challenges
  - Negation by Exception
  - Discontinuous spans
  - Reference to Definition

Thank You