Contents

1. Recap from last time

Homophonic Substitution Ciphers

- 1:>0 mappings
- English without spaces & punctuation

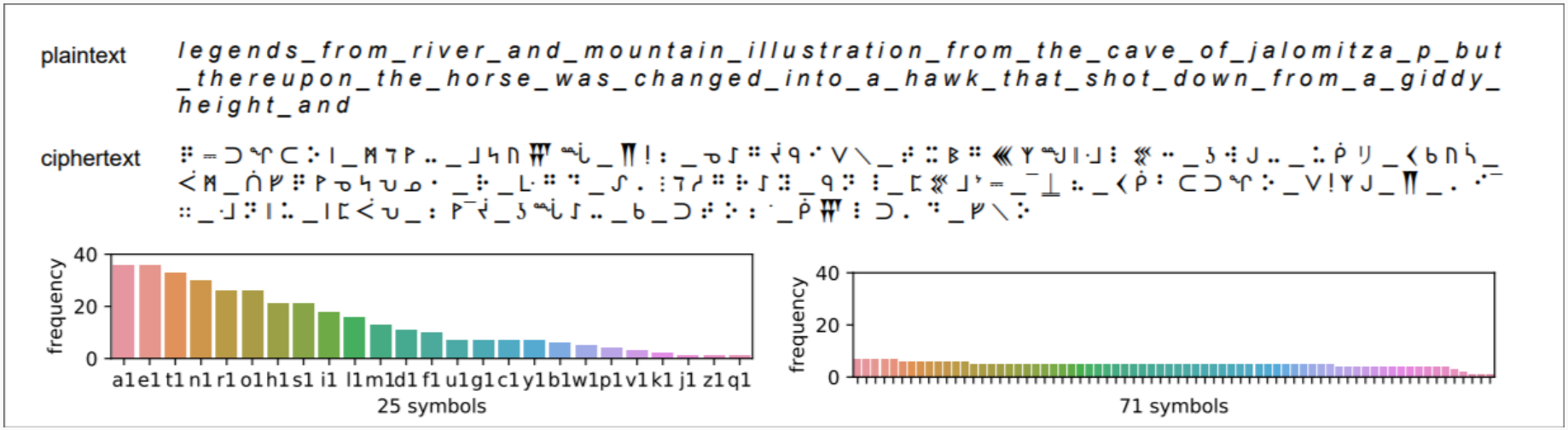

Figure 1: Example of a homophonic substitution cipher. (Kambhatla et al., Findings 2023)

1. Recap from last time

Causal LM

- Learns both cipher & plaintext
- Reads left to right

Seq2seq

- Learns plaintext only
- Bidirectional

Both suffer from $O(N^2)$ attention 😔

#keys	Model	Max Len. 400	700
30-45	Seq-to-Seq	72.30	fail
	PrefixLM	54.73	69.50
	CausalLM (tgt)	29.99	37.20
	CausalLM	**0.40**	**0.21**
40-65	PrefixLM	69.50	54.73
	CausalLM (tgt)	29.99	37.20
	CausalLM	**0.83**	**0.80**
30-85	PrefixLM	70.52	71.82
	CausalLM (tgt)	42.05	42.69
	CausalLM	**2.25**	**2.19**

Figure 2: SER on synthetic HS ciphers. (Kambhatla et al., Findings 2023)

2. Standard Attention Computation

2. Standard Attention Computation

Rows (Queries)
- Token we are looking for

Columns (Keys)
- Token we are looking at

Values (Cells)
- Attention Score

Notice: $(N \times N) = N^2$

Cipher: X Y Z

$$X \rightarrow [Y, Z]$$

$$Y \rightarrow [X, Z]$$

$$Z \rightarrow [X, Y]$$

	X	**Y**	**Z**
X	$X \rightarrow X$	$X \rightarrow Y$	$X \rightarrow Z$
Y	$Y \rightarrow X$	$Y \rightarrow Y$	$Y \rightarrow Z$
Z	$Z \rightarrow X$	$Z \rightarrow Y$	$Z \rightarrow Z$

Table 1: Attention Weight Matrix $(N \times N)$

3. Questions?