Attacking the Vigenere Cipher with Pattern Theory

Daniel Yao

Johns Hopkins University

25 April 2025

Introduction

Example

plaintext	z	e	b	r	a
key	b	a	b	a	b
ciphertext	a	e	c	r	b

Table

	a	b	c	d	e	f	g	h	i	j	k	l	m
a	a	b	c	d	e	f	g	h	i	j	k	l	m
b	b	c	d	e	f	g	h	i	j	k	l	m	n

	n	o	p	q	r	s	t	u	v	w	x	y	z
a	n	o	p	q	r	s	t	u	v	w	x	y	z
b	o	p	q	r	s	t	u	v	w	x	y	z	a

Introduction

Vigenere Cipher [5]

- A Vigenere cipher with key $k = (k_1, \ldots, k_K)$ is a map $f : A^N \to A^N$ where for $i = 1, \ldots, N$,

$$f_i(x_i) = (x_i + k_{i \bmod K}) \bmod |A|.$$

- To decipher, we need the inverse function

$$f_i^{-1}(y_i) = (y_i - k_{i \bmod K}) \bmod |A|.$$

Example

plaintext	z	e	b	r	a
key	b	a	b	a	b
ciphertext	a	e	c	r	b

Introduction

Decipher with key length $K = 3$

- iubwbu ujeagppehaqfademkegbiubwbu ujeagppehaqfakndteewljvy

Introduction

Decipher with key length $K = 3$

- iubwbu ujeagppehaqfademkegbiubwbu ujeagppehaqfakndteewljvy
- iubwbu ujeagppehaqfademkegbiubwbu ujeagppehaqfakndteewljvy

Introduction

Decipher with key length $K = 3$

- iubwbu ujeagppehaqfademkegbiubwbu ujeagppehaqfakndteewljvy
- iubwbu ujeagppehaqfademkegbiubwbu ujeagppehaqfakndteewljvy
- it was the epoch of belief it was the epoch of incredulity

Introduction

Decipher with key length $K = 3$

- iubwbu ujeagppehaqfademkegbiubwbu ujeagppehaqfakndteewljvy
- iubwbu ujeagppehaqfademkegbiubwbu ujeagppehaqfakndteewljvy
- it was the epoch of belief it was the epoch of incredulity

Solution

- abc

Introduction

Language Model [4]

- English is a stationary ergodic stochastic process with state space (alphabet)

$$A = \{a, \ldots, z, \_\}.$$

- The *character 2-gram model* is that English is a Markov chain $X_1, X_2, \ldots$ with state space A and transition matrix

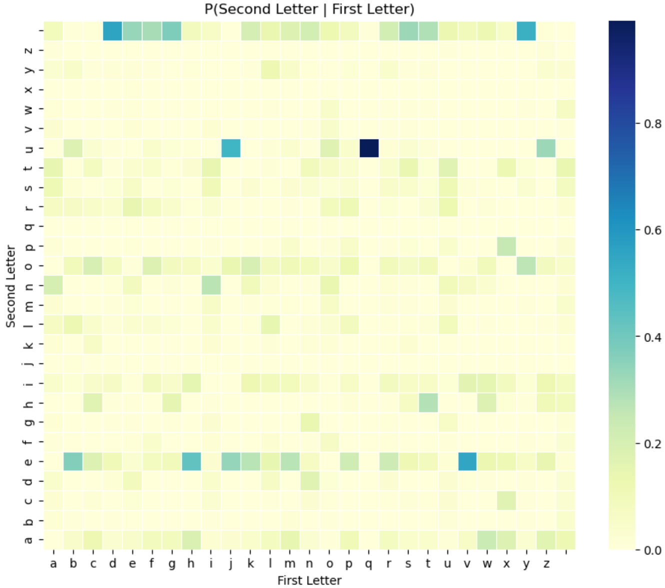
$$Q^{(2)}(x_1, x_0) = P(X_1 = x_1 \mid X_0 = x_0).$$

Example

- If $n = 2$, then $(X_0, X_1, X_2) = (\mathrm{t}, \mathrm{h}, \mathrm{e})$ has probability

$$\begin{aligned} P(\mathrm{t}, \mathrm{h}, \mathrm{e}) &= P(\mathrm{t})P(\mathrm{h} \mid \mathrm{t})P(\mathrm{e} \mid \mathrm{h}) \\ &= 0.07498(0.2828)(0.4321). \end{aligned}$$

Introduction



Methods

Maximum Likelihood Estimation

- Given an encoded text $(y_0, \ldots, y_N)$ and key length K, we want to find the decoded text $f^{-1}(y_0, \ldots, y_N)$.
- The posterior probability that f is the true cipher is

$$P(f \mid y) = \frac{P(y \mid f)P(f)}{P(y)} \propto L(f).$$

- The *likelihood function* is

$$\begin{aligned} L(f) &= P(f_0^{-1}(y_0), \ldots, f_N^{-1}(y_N)) \\ &= Q^{(1)}(f_0^{-1}(y_0)) \prod_{i=1}^{N} Q^{(2)}(f_{i-1}^{-1}(y_{i-1}), f_i^{-1}(y_i)). \end{aligned}$$

- The *objective* is to find the best cipher

$$f^* = \arg\max L(f).$$

Methods

Gibbs Sampling [4]

- The number of ciphers is $|A|^K$, so the posterior is intractable.
- Turns out, we can use a Gibbs sampling, a *Markov chain Monte Carlo* method that samples from the likelihood distribution $L(f)$.
- We want to construct a Markov chain on the set of ciphers that has the (limiting) *stationary distribution* $p_\beta(f)$ where

$$\arg\max p_\beta(f) = \arg\max L(f).$$

- The inverse temperature β controls the amount of exploration.

Methods

Experiment [1] [2] [3]

- Corpus: *Crime and Punishment* by Fyodor Dostoevsky
- Text: "It was the best of times, it was the worst of times..." ($N = 592$) from *A Tale of Two Cities* by Charles Dickens
- Key lengths: $K = 592, 296, 197, 148, 118, \ldots, 1$.
- Hyperparameters: $M = 10^5$, $\beta = 0.5$.

Results

$K = 592$

- rn cte om_ghycufot ad_ali_hothtsicute_dse_u_llno t

$K = 296$

- er wassap_o_erhtast_irge in eeduthedo_nat g_hagrki

$K = 197$

- it wat the besee_fi_widen wan_withe dr_st pe ta_ar

$K = 148$

- it was the best of te_rs it was ehe worsttof t_yes

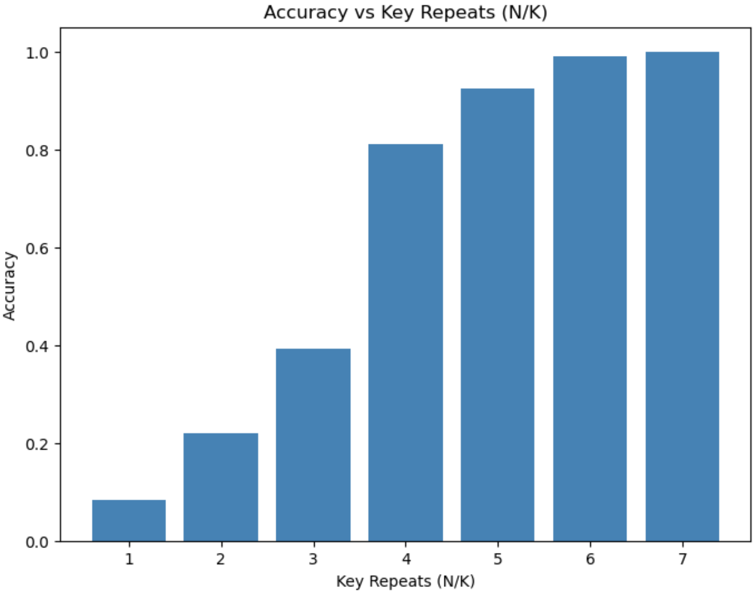
$K = 118$

- it was the best of times it was the p_rst of times

Results

K	N/K	Count	Accuracy	Run Time (s)
592	$N/1$	49	0.0828	598
296	$N/2$	130	0.2196	360
197	$N/3$	232	0.3919	303
148	$N/4$	480	0.8108	258
118	$N/5$	547	0.9240	190
98	$N/6$	586	0.9899	119
84	$N/7$	592	1.0000	76
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
10	$N/59$	592	1.0000	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	$N/592$	592	1.0000	0

Results



Accuracy vs Key Repeats (N/K)
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
Key Repeats (N/K)

Conclusion

Conclusion

- Accuracy increases with number of key repeats.
- With $N/K = 4$ key repeats, Gibbs sampling yields human-readable text.

Next Steps

- Try longer/shorter texts.
- Try stronger language model with longer n-grams.
- Try more complex ciphers (e.g., running key cipher).

References

[1] Bird, S., Loper, E., and Klein, E. (2009). *Natural Language Processing with Python*.

[2] Dickens, C. (1859). *A Tale of Two Cities*. Project Gutenberg.

[3] Dostoevsky, F. (1866). *Crime and Punishment*. Project Gutenberg.

[4] Menon, G. (2020). Pattern Theory: Old and New. Lecture notes, Brown University.

[5] Vigenère, B. d. (1586). *Traicté des Chiffres, ou Secretes Manières d'Écrire [Treatise on ciphers, or secret ways of writing]*. Abel l'Angelier.

Appendix

Gibbs Sampling [4]

- Gibbs sampling as a Markov chain Monte Carlo method to sample from the posterior distribution of σ.
- The Gibbs distribution with inverse temperature β has the pmf

$$p_\beta(x) = \frac{\exp(-\beta E(x))}{Z_\beta}$$

  where E is the energy function and Z_β is the partition function (normalization constant).
- Let $E(\sigma) = -\log L(\sigma)$. We want to find

$$\begin{aligned} \arg\max_\sigma L(\sigma) &= \mathit{arg}\max_\sigma \exp(-\beta E(\sigma)) \\ &= \arg\max_\sigma p_\beta(\sigma). \end{aligned}$$

Appendix

Gibbs Sampling [4]

procedure MCMC(E, newPerm, β, N)
  $\sigma \leftarrow \text{id}$, $\sigma^* \leftarrow \sigma$
  for $i = 1$ to N **do**
    $\tau \leftarrow \text{newPerm}(\sigma)$
    if $E(\tau) < E(\sigma)$ **or** $\text{unif}(0, 1) < \exp(-\beta \Delta E)$ **then**
      $\sigma \leftarrow \tau$
    end if
    if (**then**$E(\sigma) < E(\sigma^*)$)
      $\sigma^* \leftarrow \sigma$
    end if
  end for
  return σ^*
end procedure