Neural Cryptanalysis of Classical Ciphers

https://youtu.be/_Vc-oZ6U3CA

HSU 한성대학교 HANSUNG UNIVERSITY

CryptoCraft LAB
https://crypto.modoo.at

Contents

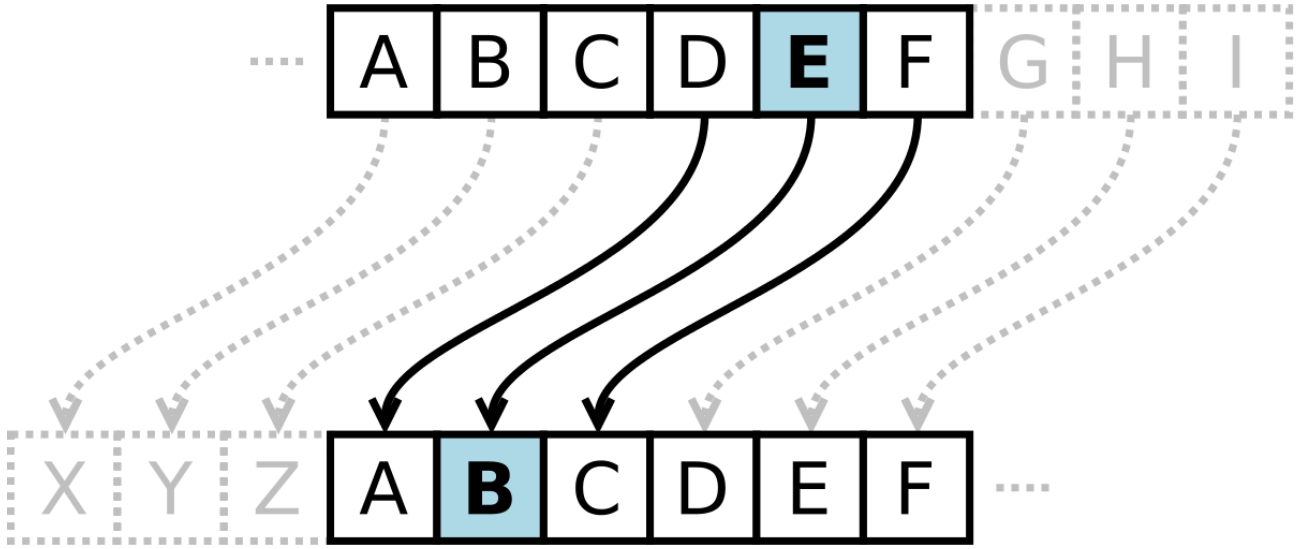
CryptoCraft LAB

개요

- 인공 신경망을 이용하여 암호 취약점 분석을 지원하는 것이 목표
- 간단한 고전 암호들을 바탕으로 이를 구현
- 평문과 암호문을 통해서 키값을 복구
- 카이사르 암호, 비즈네르 암호, 치환 암호

카이사르 암호

- 주어진 키값에 따라 알파벳을 미는 것
- 시저가 비밀리에 편지를 보낼 때 쓰던 암호



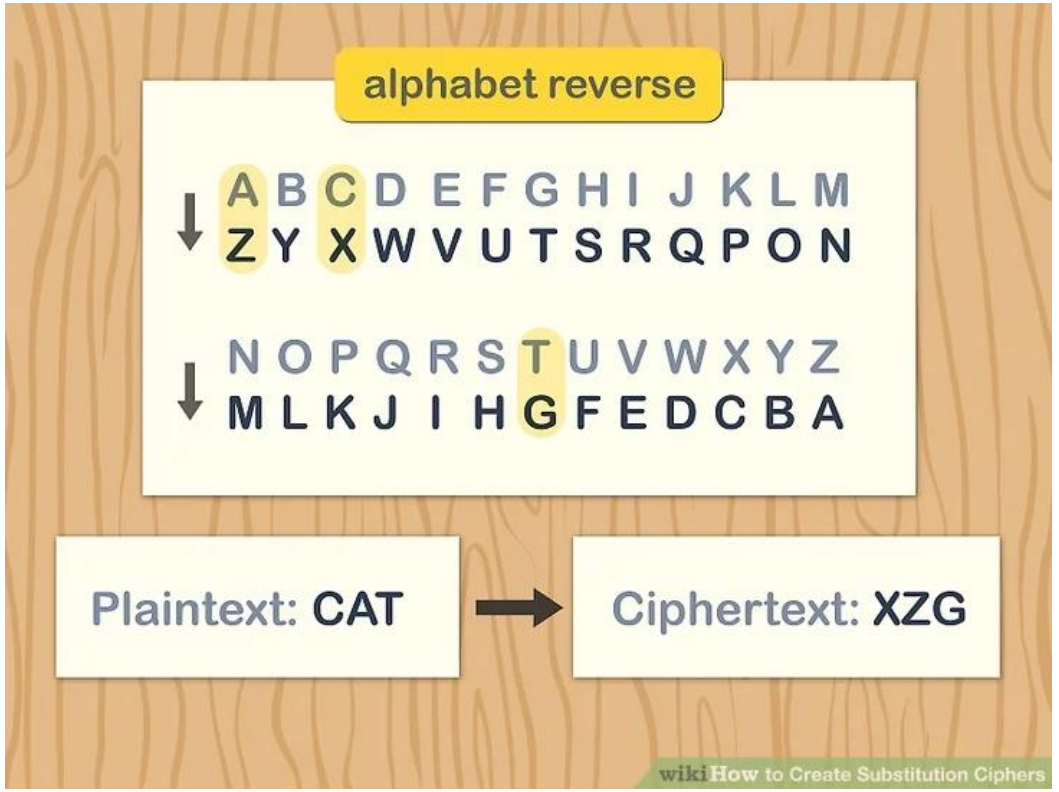
비즈네르 암호

- 카이사르 암호와 유사하나, 키값이 단어 혹은 문장
- 키값을 이용하여 암호문 구함

치환 암호

- 알파벳 순서를 섞은 뒤 치환

- 키에 따라 다양한 결과값 생성 가능



관련 연구

- RNN을 이용한 에니그마 학습
- 최소 백만 번의 훈련이 필요
- K40 GPU로 며칠씩 학습 필요
- 96~97%의 정확성

관련 연구

- CipherGAN
- GAN (Generative Adversarial Networks)
- 쉬프트 암호와 비즈네르 암호 공격 성공
- 쉬프트 암호는 98.7%, 비즈네르 암호는 75.7%의 정확성 얻음

배경

- P : 평문 (plaintexts)
- C : 암호문 (ciphertexts)
- K : 키 (keys)
- Z_{26} : 모든 연산이 26개의 알파벳 내에서 일어남

정의 1

- 평문 x ∈ P
- 암호문 y ∈ C
- 키 k ∈ K
- P = C = K = Z_{26}

- 카이사르 암호화는 $E_k(x) = x + k \% 26$
- 카이사르 복호화는 $D_k(y) = y - k \% 26$

정의 2

- 평문 x ∈ P
- 암호문 y ∈ C
- 키 k ∈ K
- P = C = K = Z_{26} && $Z_{26}^m = Z_{26} * Z_{26} * \dots * Z_{26}$ (m번 반복)

- K = $(k_1, \dots, k_m)$
- 비즈네르 암호화는 $E_{k_1,\dots,k_m}(x_1,\dots x_m) = (x_1 + k_1,\dots, x_m + k_m)$ % 26
- 비즈네르 복호화는 $D_{k_1,\dots,k_m}(y_1,\dots y_m) = (y_1 - k_1,\dots, y_m - k_m)$ % 26

정의 3

- 평문 x ∈ P
- 암호문 y ∈ C
- 키 ρ ∈ K
- P = C = Z_{26} && K = { ρ | ρ는 0, 1, … , 25의 순열 }

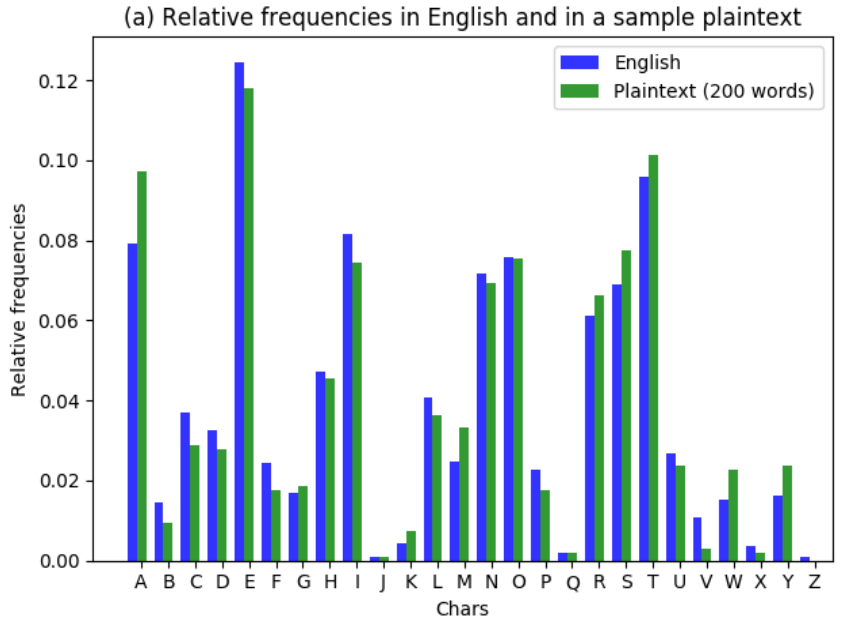
- 치환암호 암호화는 $E_\rho(x) = \rho(x)$
- 치환암호 복호화는 $D_\rho(y) = \rho^{-1}(x)$

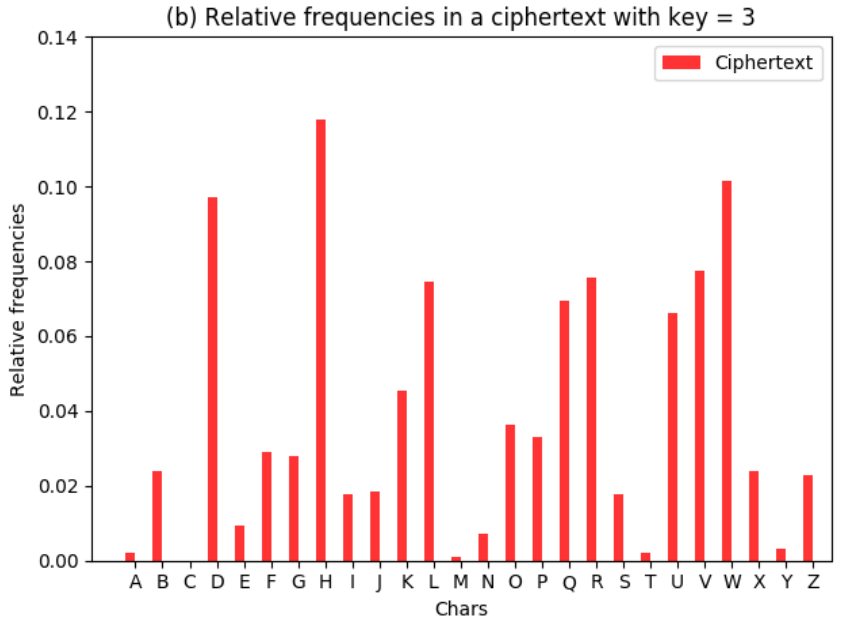
Experimental setup

- BNC(British National Corpus)에서 백만 개의 말뭉치 추출
- 신경망은 케라스 라이브러리와 텐서플로우를 이용하여 구현
- 맥북 프로 Intel Core i7 2GHz, 16GB RAM
- GPU나 멀티코어 최적화 없음

카이사르 암호

- 평문과 암호문에서의 글자 빈도수
- 키 = 3
- 히스토그램의 유사성을 통해 브루트-포스 공격이 굳이 필요하지 않다는 것을 알 수 있음





카이사르 암호

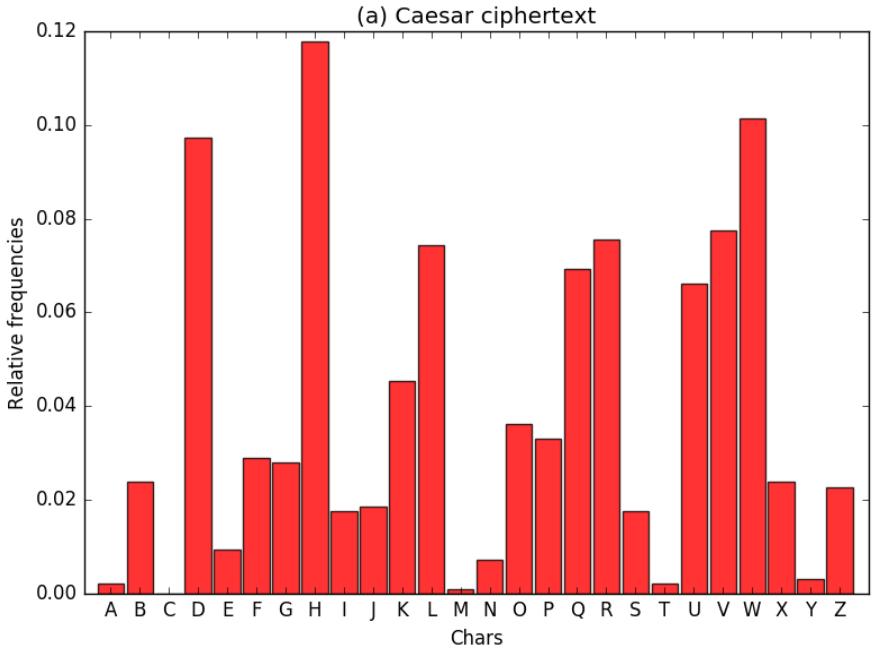
1. 랜덤 키를 통한 암호화된 평문 데이터를 많이 준비함
2. 암호문에서의 글자별 빈도수를 계산함
3. 주어진 빈도수를 통해 신경망을 학습시키고, 일치하는 정도가 키(아웃풋)가 됨
4. 독립된 다른 데이터셋에서 테스트

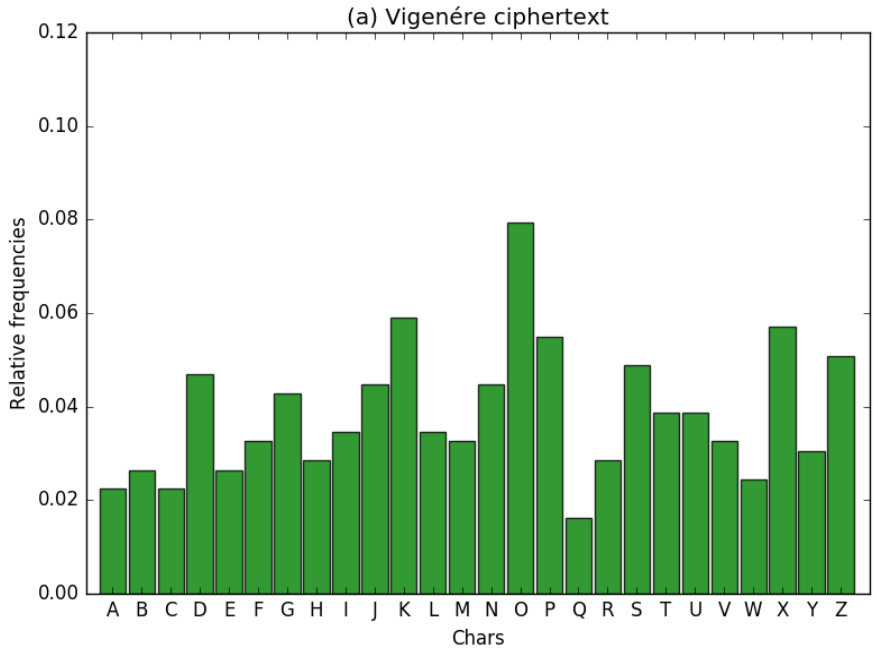
카이사르 암호

- 100개 단어를 각기 다른 키로 암호화 한 5097개의 암호문을 통해서 학습
- 훈련은 30초 정도 걸렸고, 해독은 거의 즉각적
- 30 단어까지는 매우 정확

비즈네르 암호

- 카이사르 암호문과 비즈네르 암호문의 글자 빈도수 비교





비즈네르 암호

- 길이 m의 패스워드를 길이 m의 각 평문 블록에 추가
- m을 구하기 위해 브루트-포스 공격 필요
- 카이사르 암호를 깰 때 이용하였던 인공지능 재사용

비즈네르 암호

- C = $c_0, c_1, \ldots, c_{n-1}$
- MAX = 공격자가 시도하려는 키의 최대 길이

- integer m of [1, MAX]:
- C의 모든 하위 텍스트 S_i는 거리 m의 글자들로 이루어져 있음
- ex. $c_0, c_m, c_{2m}, \ldots, c_1, c_{m+1}, c_{2m+1}, \ldots, c_{m-1}, c_{2m-1}, c_{3m-1}, \ldots;$
- 카이사르 암호 분류기에 모든 S_i의 s 적용
- 분류기가 모든 S_i의 s에 대한 한계보다 더 큰 값을 리턴하게 되면, 키를 찾은 것

- 이 공격의 복잡도는 O(MAX^2)

치환 암호

- 200 단어의 평문 (스페이스 제거 및 뒷부분 생략)

  MAYWANTTOMODIFYSUCHFORMULATIONSALONGTHELINESAREADERISLIKEL
  YTOFEELATTHISPOINTORMANYREADERSMAYRESPONDBYORPOSSIBLYBETTE
  RBYTACKLINGTHEISSUEOFHETEROGENEOUSREADERRESPONSESMOREDI…

- 치환암호 키 : VETISLFMBDGNCYQHJPXZAORKUW

  CVURVYZZQCQIBLUXATMLQPCANVZBQYXVNQYFZMSNBYSXVPSVISPBXNBGSN
  UZQLSSNVZZMBXHQBYZQPCVYUPSVISPXCVUPSXHQYIEUQPHQXXBENUESZZS
  PEUZVTGNBYFZMSBXXASQLMSZSPQFSYSQAXPSVISPPSXHQYXSXCQPSIB…

- 영어의 상대적 빈도수에 따라 암호문으로부터 매핑할 경우, 5개의 글자들(B,E,H,K,V)만 일치

  GOFYORIITGTLAPFNCUHPTSGCDOIATRNODTRWIHEDARENOSEOLESANDAKED
  FITPEEDOIIHANMTARITSGORFSEOLESNGOFSENMTRLBFTSMTNNABDFBEIIE
  SBFIOUKDARWIHEANNCETPHEIESTWERETCNSEOLESSENMTRNENGTSELA…

치환 암호

- n-gram을 이용하여 주어진 텍스트가 영어 텍스트에서 얼마나 떨어져 있는지를 평가
- 두 글자를 바꾸어 더 나은 치환을 찾음
- 신경망을 통해 텍스트를 평가

치환 암호

1. 랜덤 키에 의해 암호화 된 충분한 양의 영어 평문 데이터셋을 구함

2. 평문과 암호문의 3-gram의 빈도수를 계산

3. 빈도수를 인풋으로 하여 신경망을 학습시킨 뒤, 평문일 경우 1, 암호문일 경우 0인 1비트 데이터를 아웃풋으로 설정

4. 독립된 데이터셋에서 신경망을 테스트

치환 암호 - 공격 전략

1. 랜덤 키(key_i)를 선택

2. 신경망을 이용하여 $goodness_i$ 값을 구함

3. For MAXSWAPS (가능한 스왑 최대치) iterations:
- (a) 두 글자를 스왑하고 $goodness_i$를 다시 계산
- (b) 만약 $goodness_i$ 값이 더 나을 경우 새로운 키를 key_i에 저장

치환 암호

- 아까 얻었던 (불만족스러운) 복호화 된 텍스트

  GOFYORIITGTLAPFNCUHPTSGCDOIATRNODTRWIHEDARENOSEOLESANDAKED
  FITPEEDOIIHANMTARITSGORFSEOLESNGOFSENMTRLBFTSMTNNABDFBEIIE
  SBFIOUKDARWIHEANNCETPHEIESTWERETCNSEOLESSENMTRNENGTSELA…

- 신경망이 goodness 값을 0.38로 평가, C와 H를 바꿀 경우 goodness 값 = 0.78

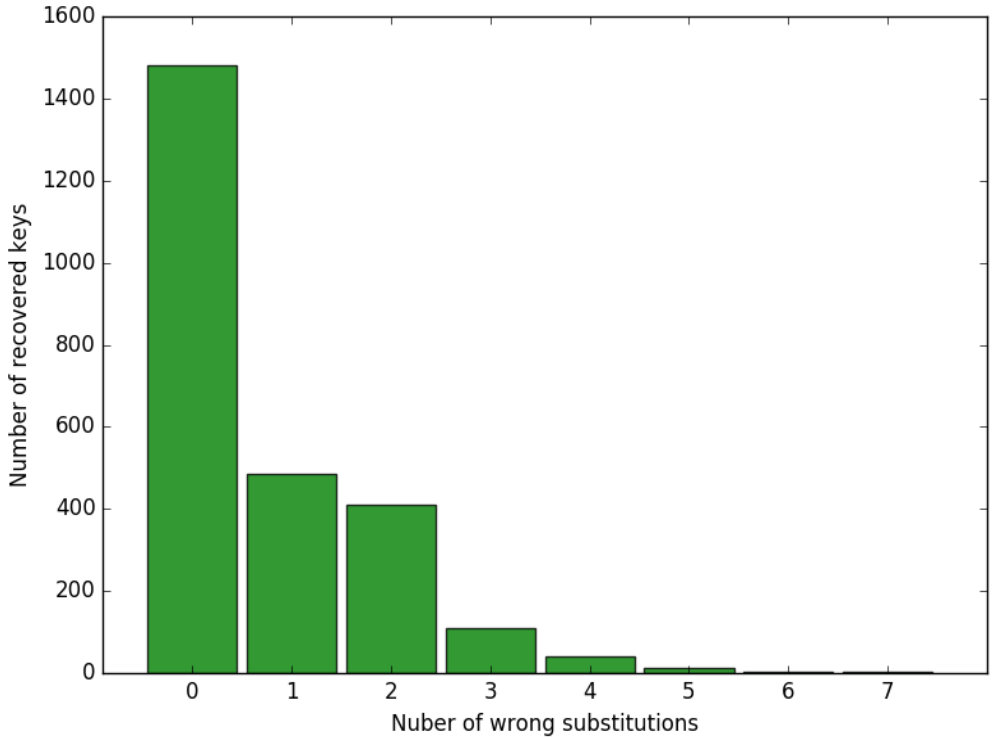
  GOFYORIITGTLAPFNHUCPTSGHDOIATRNODTRWICEDARENOSEOLESANDAKED
  FITPEEDOIICANMTARITSGORFSEOLESNGOFSENMTRLBFTSMTNNABDFBEIIE
  SBFIOUKDARWICEANNHETPCEIESTWERETHNSEOLESSENMTRNENGTSELA…

- goodness 값 계속 교환하면서 0.9998의 goodness 값을 통해 복호화한 결과:

  MAYWANTTOMODIFYSUCHFORMULATIONSALONGTHELINESAREADERISLIKEL
  YTOFEELATTHISPOINTORMANYREADERSMAYRESPONDBYORPOSSIBLYBETTE
  RBYTACKLINGTHEISSUEOFHETEROGENEOUSREADERRESPONSESMOREDI…

치환 암호

- 200개 단어 2500개 글자로 테스트
- 58%의 키 완전 복구 (1450 글자)
- 93%의 케이스에서 2개 이하의 오류



Q & A