Report za zimný semester

Úvod

Tento projekt sa zaoberá problémom opätovného použitia jednorazového kľúča, v kryptografii známym ako *Two-time pad*. Východiskom bola metodológia opísaná vo vedeckej publikácii *Mason et al. (CCS 2006)* [1], kde autori úspešne rekonštruovali textové správy zo sčítaných šifrových textov pomocou jazykových modelov.

Mojím cieľom bolo adaptovať tento prístup na obrazové dáta. Kým v texte sa využíva pravdepodobnosť nasledujúceho znaku, pri obraze som pracoval s pravdepodobnosťou hodnoty pixelu na základe jeho lokálneho okolia.

Implementácia

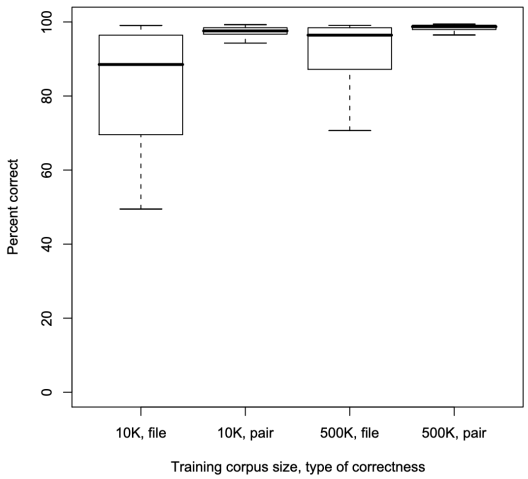
Implementácia bola realizovaná v jazyku **Python**. Na maticové operácie a efektívny výpočet pravdepodobností som využil knižnicu **NumPy**, pre spracovanie obrazových dát knižnicu **OpenCV**.

Replikácia na texte

Pred samotnou aplikáciou na obrazy som sa zameral na overenie teoretických poznatkov replikáciou pôvodného experimentu na textových dátach. Ako trénovací korpus pre vytvorenie jazykového modelu som využil texty z verejne dostupných digitálnych kníh.

V modeli som použil kontext o veľkosti 5 znakov – pre každú sekvenciu piatich znakov bola vypočítaná pravdepodobnosť výskytu nasledujúceho znaku.

Dosiahnutá úspešnosť rekonštrukcie celého súboru (správne priradenie znaku do konkrétneho súboru) bola 62 %. Úspešnosť rekonštrukcie párov (správne dešifrovaný znak, aj v prípade zámeny medzi súbormi) dosiahla 95 %. Tieto výsledky sú porovnateľné s dátami prezentovanými v pôvodnej vedeckej práci.



Obr. 1: Úspešnosti z Mason et al. [1]

Experimenty s obrazovými dátami MNIST

Pre experimenty s obrazovými dátami som zvolil dataset MNIST, ktorý pozostáva z rukou písaných číslic. Obrazy sú v odtieňoch sivej (*grayscale*) s fixným rozlíšením 28×28 pixelov. Ako trénovací korpus pre vytvorenie pravdepodobnostného modelu som použil vzorku 20 000 obrazov. Cieľom bolo určiť hodnotu pixelu x na základe hodnôt v jeho okolí.

Metodika vyhodnocovania

Na posúdenie kvality rekonštrukcie som implementoval vyhodnocovací skript, ktorý porovnáva dvojicu výsledných obrazov (D_1, D_2) s originálmi (O_1, O_2).

Priemerná absolútna chyba (MAE)

Základnou metrikou je priemerná absolútna chyba, ktorou porovnávame dva obrazy O a D. Pre obrazy s rozmermi $M \times K$ je definovaná ako:

$$\mathrm{MAE}(O, D) = \frac{1}{M \cdot K} \sum_{i=1}^{M} \sum_{j=1}^{K} |p_{O,i,j} - p_{D,i,j}|$$

kde $p_{O,i,j}$ a $p_{D,i,j}$ sú hodnoty jasu pixelov na rovnakom mieste v oboch obrazoch.

Párovanie obrazov

Keďže model generuje dva obrazy bez určenia ich poradia, musíme najskôr identifikovať, ktorý výsledok patrí ku ktorému originálu. Porovnávame dve možné kombinácie:

- **Priama:** O_1 s D_1 a O_2 s D_2,
- **Krížová:** O_1 s D_2 a O_2 s D_1.

Finálne priradenie zodpovedá tej kombinácii, ktorá minimalizuje súčet hodnôt MAE.

Definícia výsledných metrík

Po určení korektných párov počítam dve hlavné veličiny:

- **Priemerná chyba (E):** Je to priemerná hodnota MAE vypočítaná z oboch spárovaných obrazov. Táto veličina udáva priemerný rozdiel jasu.
- **Priemerná úspešnosť (U):** Podiel pixelov, ktorých jas sa od originálu líši o menej ako stanovenú toleranciu $T = 10$. Najskôr definujeme funkciu $S(i, j)$:

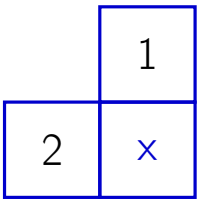
$$S(i,j) = \begin{cases} 1, & \text{ak } |p_{O,i,j} - p_{D,i,j}| \leq T \\ 0, & \text{ak } |p_{O,i,j} - p_{D,i,j}| > T \end{cases}$$

$$U = \frac{1}{M \cdot K} \sum_{i=1}^{M} \sum_{j=1}^{K} S(i,j)$$

1. Model so základným kontextom

V prvom experimente som využil jednoduchý kontext, ktorý bral do úvahy iba bezprostredných susedov aktuálne spracovávaného pixelu. Kontext tvorili:

- 1 pixel vľavo,
- 1 pixel hore.



Obr. 2: Vizualizácia základného kontextu pre pixel x

Dosiahnuté výsledky pre tento model:

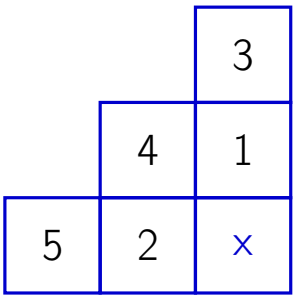
- Priemerná chyba E: 31,8 odtieňov.
- Priemerná úspešnosť U: 81,0%

2. Model s rozšíreným kontextom

V druhom experimente som rozšíril sledované okolie cieľového pixelu x. Kontext tvorili:

- 2 pixely vľavo,
- 2 pixely hore,
- 1 pixel vľavo hore (diagonálne).

V prípadoch, kde tento rozšírený kontext nebol dostupný, som použil model z predchádzajúcej sekcie 2.1.



Obr. 3: Vizualizácia rozšíreného kontextu pre pixel x

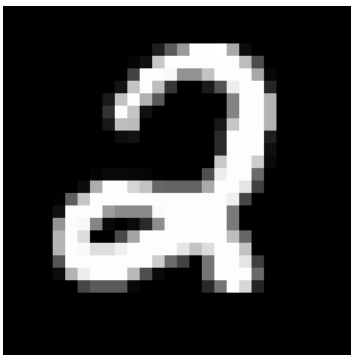
Dosiahnuté výsledky pre tento model:

- Priemerná chyba E: 33,3 odtieňov.
- Priemerná úspešnosť U: 81,0%

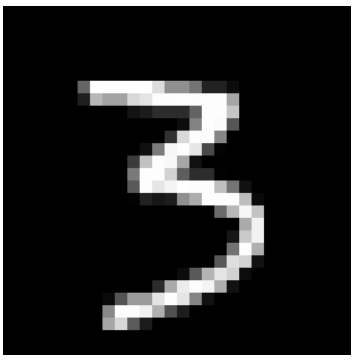
Analýza chýb

Pri vyhodnocovaní rekonštruovaných obrazov som narazil na špecifický typ chyby, ktorý sa v numerických štatistikách stráca.

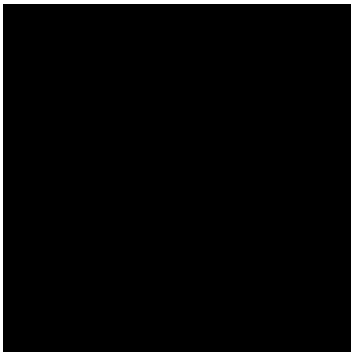
Namiesto separácie model priradil takmer všetku obrazovú informáciu do jedného výstupu a druhý výstup nechal takmer kompletne čierny.



(a) Originál 1



(b) Originál 2



(c) Rekonštrukcia 1 (Chyba: čierny obraz)

(d) Rekonštrukcia 2 (Chyba: zmes)

Obr. 4: Ukážka neúspešnej separácie. Model neoddelil číslice, ale vytvoril jeden prázdny obraz a jeden zmiešaný obraz.

Zhodnotenie a budúca práca

Vysoká priemerná úspešnosť je do veľkej miery ovplyvnená charakterom datasetu MNIST. Tieto obrazy obsahujú veľké množstvo čierneho pozadia, najmä na okrajoch. Model dokáže veľmi ľahko uhádnuť tieto čierne pixely, čím si umelo zvyšuje celkovú štatistickú úspešnosť, aj keď v kritickej oblasti (samotná číslica) zlyháva, ako vidno na ukážke vyššie.

V letnom semestri sa preto plánujem zamerať na odstránenie tohto nedostatku:

1. **Vlastný dataset:** Vytvorím sadu vlastných fotografií, ktoré nebudú mať uniformné čierne pozadie a budú ukazovať menšiu vzájomnú podobnosť ako číslice MNIST.

2. **Vylepšenie modelu:** Aktuálny model naráža na problém príliš veľkého počtu kombinácií. Pri 256 odtieňoch a viacerých susedoch existuje priveľa možností, ktoré model v tréningu nikdy nevidel. Plánujem preto zredukovať počet farieb zhlukovaním podobných odtieňov do skupín.

Literatúra

[1] J. Mason, K. Small, A. Stubblefield, and D. Wallach. *A Natural Language Approach to Automated Cryptanalysis of Two-time Pads.* In Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS '06), 2006.