

Katedra informatiky
Přírodovědecká fakulta
Univerzita Palackého v Olomouci

DIPLOMOVÁ PRÁCE

Prolamování CAPTCHA zabezpečení



2019

Bc. Kamil Hanus

Vedoucí práce: Mgr. Martin
Trnečka, Ph.D.

Studijní obor: Informatika, prezenční
forma

Bibliografické údaje

Autor:	Bc. Kamil Hanus
Název práce:	Prolamování CAPTCHA zabezpečení
Typ práce:	diplomová práce
Pracoviště:	Katedra informatiky, Přírodovědecká fakulta, Univerzita Palackého v Olomouci
Rok obhajoby:	2019
Studijní obor:	Informatika, prezenční forma
Vedoucí práce:	Mgr. Martin Trnečka, Ph.D.
Počet stran:	23
Přílohy:	1 CD/DVD
Jazyk práce:	český

Bibliographic info

Author:	Bc. Kamil Hanus
Title:	Breaking the CAPTCHA
Thesis type:	master thesis
Department:	Department of Computer Science, Faculty of Science, Palacký University Olomouc
Year of defense:	2019
Study field:	Computer Science, full-time form
Supervisor:	Mgr. Martin Trnečka, Ph.D.
Page count:	23
Supplements:	1 CD/DVD
Thesis language:	Czech

Anotace

Ukázkový text závěrečné práce na Katedře informatiky Přírodovědecké fakulty Univerzity Palackého v Olomouci, který je zároveň dokumentací stylu pro text práce v \LaTeX . Zdrojový text v \LaTeX je doporučeno použít jako šablonu pro text skutečné závěrečné práce studenta.

Synopsis

Sample text of thesis at the Department of Computer Science, Faculty of Science, Palacký University Olomouc and, at the same time, documentation of the \LaTeX style for the text. The source text in \LaTeX is recommended to be used as a template for real student's thesis text.

Klíčová slova: styl textu; závěrečná práce; dokumentace; ukázkový text

Keywords: text style; thesis; documentation; sample text

Děkuji, děkuji, děkuji.

Místopřísežně prohlašuji, že jsem celou práci včetně příloh vypracoval/a samostatně a za použití pouze zdrojů citovaných v textu práce a uvedených v seznamu literatury.

datum odevzdání práce

podpis autora

Obsah

1	Úvod	7
2	Captcha	8
2.1	Historie	8
2.2	Druhy kódů	8
2.3	Možnosti prolamování	10
2.3.1	Strojové učení	10
3	Captcha Breaker	12
3.1	Použité technologie	12
3.2	Prolamované typy	14
3.3	Extrakce symbolů	14
3.3.1	Podporované grafické operace	14
3.4	Uživatelská část	14
3.5	Administrační rozhraní	15
	Závěr	19
	Conclusions	20
A	První příloha	21
B	Druhá příloha	21
C	Obsah přiloženého CD/DVD	21
	Seznam zkratk	23

Seznam obrázků

- | | | |
|---|---|----|
| 1 | Příklad obrázku, které generovala služba reCAPTCHA (verze 1) obsahující text „Chomsky fake“ | 9 |
| 2 | Varianta reCAPTCHA v2, kde má uživatel vybrat všechny dlaždice obsahující semafor. | 11 |
| 3 | Varianta GeeTest CAPTCHA používaná na webu binance.com | 12 |

Seznam tabulek

- | | | |
|---|--|----|
| 1 | URL endpointy pro globální namespace | 15 |
| 2 | URL endpointy pro namespace /admin/ | 16 |

Seznam vět

Seznam zdrojových kódů

1 Úvod

V posledních letech urazila technologie strojového učení a umělé inteligence velký posun vpřed. Ačkoliv historie vzniku termínu strojové učení sahá do 60. let 20. století, masová adaptace nastává až v posledních několika letech, kdy se lze s pojmy umělá inteligence či AI (z anglického *artificial intelligence* ¹. Můžeme se domnívat, že jistou spojitost s nárůstem zájmu o strojové učení má i zpřístupnění open-source nástrojů, umožňujících snadnou aplikaci příslušných algoritmů, široké veřejnosti ².

Cílem této diplomové práce je rozšíření autorových znalostí v oblasti umělé inteligence. Jako vhodný problém na kterém lze demonstrovat potenciál strojového učení se naskýtá prolamování CAPTCHA zabezpečení. Aktuálně dostupné nástroje strojového učení poskytují elegantní způsob jak prolamovat klasické CAPTCHA zabezpečení. Zřejmě proto lze pozorovat vývoj nových typů CAPTCHA zabezpečení, které využívají některé prvky umělé inteligence. Diplomovou práci lze rozdělit do dvou částí – první se zabývá samotným problémem CAPTCHA zabezpečení, jeho historií a možným prolamováním. Druhá popisuje aplikaci nazvanou CaptchaBreaker, která popisuje možnost jak lze v rozumném čase řešit problém dekodování některých typů CAPTCHA kódů.

¹https://trends.google.com/trends/explore?date=all&q=%2Fm%2F01hyh_

²https://trends.google.com/trends/explore?date=all&q=%2Fm%2F01hyh_,%2Fg%2F11gd3905v1,%2Fg%2F11bwp1s2k3,%2Fg%2F11c1r2rvnp

2 Captcha

2.1 Historie

Za akronymem CAPTCHA stojí anglický text „**C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part“. Obecně je principem tohoto testu vygenerování určitého typu dotazu a ověření odpovědi, typicky na webových stránkách. Na základě verifikace odpovědi je rozhodnuto, zda systém komunikuje s reálným uživatelem nebo robotem. Některé zdroje uvádějí [česká wiki] pouze přepis textu z deformovaného obrazu do vstupního pole, což lze považovat za zavádějící, jelikož existuje mnoho dalších variant.

Důvod pro zavedení takové ochrany je prostý. Pokud je umožněno na web zaslat libovolné požadavky bez omezení (například. příspěvky do diskuzních fór, zakládání uživatelských účtů, pokusy o přihlášení, atd.), lze očekávat že takové příležitosti využije útočník a začne systém zahlcovat, možná s úmyslem zamezit přístupnosti či šíření spamu. Ovšem je nutné mít na paměti, že nutnost ověřovat svou autentičnost je pro uživatele snížením komfortu. Proto se objevují technologie které během prvotní interakce nepožadují ověření žádné, nebo ověřují uživatele natolik sofistikovanými metodami, že není nutná jeho interakce.

Zajímavým použitím technologie byla digitalizace archivu článků deníku The New York Times či literárních děl společností Google. Slova, která nebylo možné strojově rozpoznat s dostatečnou přesností byla dekodována s použitím nástroje reCAPTCHA. Metoda spočívala vždy v zobrazení obrázků dvou slov v kódu, přičemž u jednoho z nich byl známý jeho obsah. Pokud uživatel zadal správně známou část, o druhé také uvažovalo že je zadaná správně. Samozřejmě takové řešení by bylo příliš naivní a proto lze usuzovat, že za popsáním procesem existuje ještě další ověřovací mechanismus, který porovnává odpovědi uživatelů pro konkrétní slova a přiřadí jim váhu dle jejich výskytu.

2.2 Druhy kódů

<https://dynamapper.com/blog/514-online-captcha-solving-services-and-available-captcha-types> <https://ai.google/research/pubs/pub43464> - textová, obrázková, audio, noCaptcha, aka slider na binance, přihlášení přes sociální media (bot nemá účet), honeypot (skryté fieldy, člověk nevidí, bot ano a tak je vyplní)

Leetspeak

Na úvod je vhodné zmínit tzv. *leetspeak* kódy. V raných fázích internetu se jednalo o přepis znaků, které s užitím jisté míry představivosti tvořily smysluplný text (např. *N04m* jako ekvivalent slova *Noam*), avšak pro útočníka bez znalosti generujícího mechanismu bylo obtížné automatizovat vyplňování kontrolních otázek. Takové ověření bylo možné potkat např. na některých IRC(odkaz co to je) kanálech [zdroj].



Obrázek 1: Příklad obrázku, které generovala služba reCAPTCHA (verze 1) obsahující text „Chomsky fake“.

Jednoduchá otázka

Velmi triviální metodou, která lze implementovat CAPTCHA zabezpečení, je použít jednoduchou textovou hádanku. Jedná se například o otázku typu *Jaký je dnes den?* očekávající na vstupu jednoslovný název dne v týdnu. Obdobně lze hovořit o jednoduchých matematických otázkách *Kolik je 4 krát 6?*. V případě, že není použit sofistikovaný nástroj generující nové druhy otázek je řešení tohoto zabezpečení triviální. Vyvine-li útočník dostatečné úsilí, po několika vygenerování různých otázek lze popsat gramatiku CAPTCHA schématu. Naivním řešením je zkonstruovat mapovací tabulku jednotlivých slov na funkce, se navzájem aplikují. Vzhledem ke konečnosti generovaných výrazů je však možné, že vytvoření formální gramatiky popisující jazyk CAPTCHA schématu a následné vytvoření překladače by mělo větší úspěšnost. Vzhledem k velmi nízkému rozšíření (odkaz) tohoto typu zabezpečení však lze usuzovat, že takové úsilí vyvine zanedbatelný počet útočníků. (přesunout možnost prolomení do jiné kapitoly)

Zkreslený text

Stále běžně se vyskytujícím typem CAPTCHA zabezpečení je přepis textu z deformovaného obrazu uživatelem. Předpokládá se, že bot není schopný vlivem zkreslení provést extrakci jednotlivých symbolů z obrazu. Ztížit čtení obrazu je možné mnoha způsoby - rozostřením textu, spojením jednotlivých znaků, vložením zkreslujících křivek, vynecháním malých částí znaků, atd. V další kapitole se diplomová práce zaměřuje právě na možnost prolomení typů CAPTCHA využívající zkreslování textu v obraze.

Identifikace objektů v obraze

Uživatelsky komfortnějším přístupem, jak rozlišit člověka od počítače, je zobrazit uživateli množinu různých obrázků a požadovat výběr pouze těch, které obsahují nějaký předmět. Zaměříme-li se na technologii Google reCaptcha [odkaz], lze pozorovat dva přístupy k tomuto problému. První variantou je zobrazit uživateli 9 obrázků, ze kterých je například vyžadován výběr pouze těch obsahující dopravní značky. Druhý přístup je segmentace obrazu do několika částí

a vyžadování výběru pouze těch, které obsahují určitý objekt (například auto). Využití využití podobného přístupu se také využívá k monetizaci obsahu majitelů webů, jelikož existují technologie [odkaz] používající CAPTCHA jako inzertní plochu.

Skládání puzzle

Další variantou CAPTCHA technologie pracující s obrazem nazvěme skládáním puzzle (společnost *GeeTest* [3](#) používá anglický název „slide“). Přístup spočívá v přesunu části obrazu ve tvaru dílku puzzle zpět na své původní místo pomocí posuvníku. Během procesu ověření je sledováno několik faktorů ovlivňujících vyhodnocovací proces, například doba posuvu, cesta myši, doplňky prohlížeče.

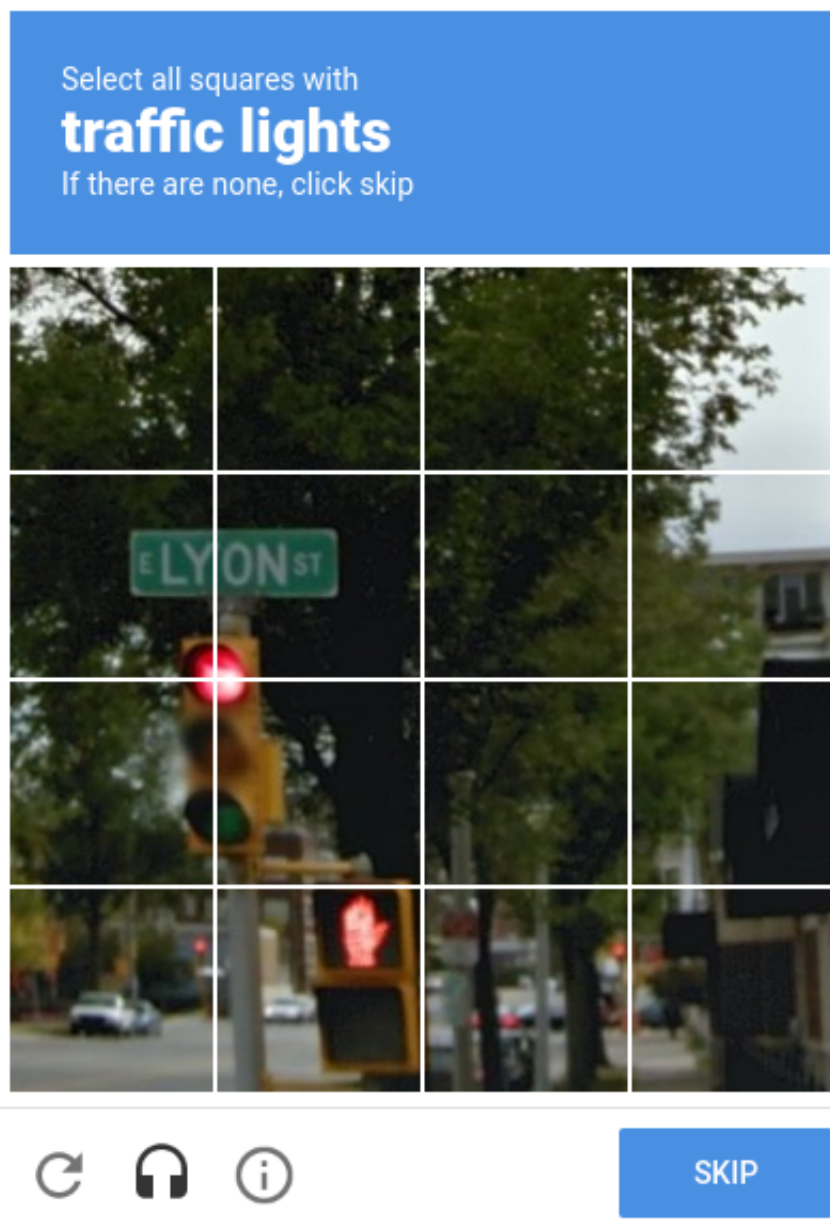
Audio CAPTCHA

Zabezpečení pomocí textové/obrazové kontrol má nevýhodu v tom, že velmi limituje uživatele se zrakovým postižením, kteří například vnímají odlišně barevné spektrum. Právě proto je vhodné jako sekundární možnost nabídnout tzv. audio CAPTCHA zabezpečení, jehož princip spočívá v přepisu znaků které jsou vloženy do audio záznamu. I zde lze najít různé typy zkreslení zvuku, které zvyšují obtížnost dekodování nahrávky.

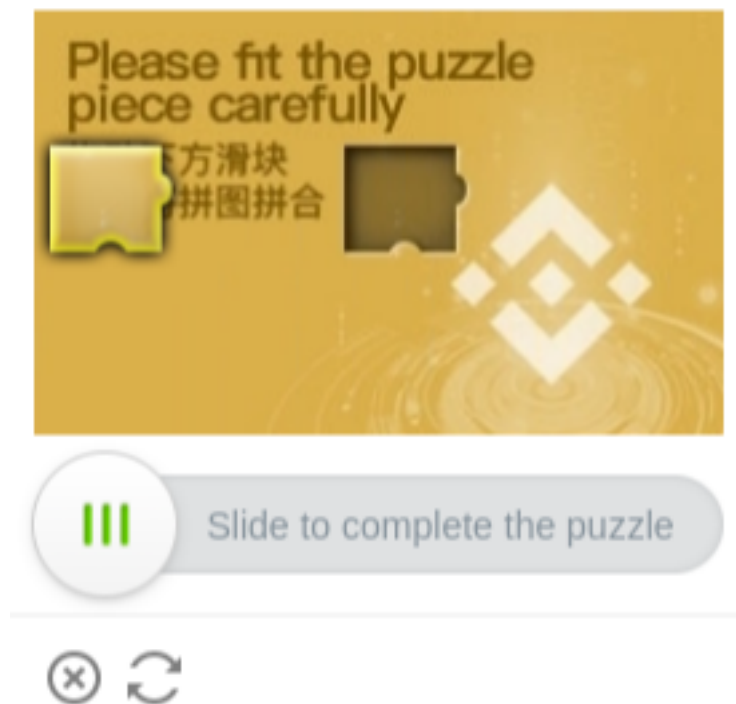
2.3 Možnosti prolamování

2.3.1 Strojové učení

- NN a CNN



Obrázek 2: Varianta reCAPTCHA v2, kde má uživatel vybrat všechny dlaždice obsahující semafor.



Obrázek 3: Varianta GeeTest CAPTCHA používaná na webu binance.com

3 Captcha Breaker

Cílem diplomové práce bylo demonstrovat možnosti prolamování CAPTCHA zabezpečení, přičemž hlavní inspirace pocházela z textu *How to Mimic Humans, Guide for Computers*. Byly vybrány různé typy zabezpečení, které ověřují rozpoznání zkráceného textu v obraze. S ohledem na odlišný postup rozpoznání segmentovaných částí obrazů byly prolamovány některé shodné typy zabezpečení.

Výsledná aplikace se skládá ze dvou částí – algoritmu rozpoznávajícího různé typy CAPTCHA zabezpečení a webové aplikace, která umožňuje pohodlnou tvorbu datasetu, klasifikátoru a umožňuje prolomit známá schémata.

3.1 Použité technologie

Python

Flask

Flask je microframework určený pro tvorbu webových aplikací napsaný v jazyce Python. Samotné jádro frameworku v základu obsahuje s nadsázkou pouze nástroje pro směrování požadavků a šablonovací systém Jinja. Vyžadujeme-li funkcionality navíc, jako například práci s DB, ORM mapování, validaci formulářů nebo autentizaci požadavků, je nutné doinstalovat příslušný modul. Framework se tedy snaží být minimalistický a je pouze na uvážení vývojáře, jaké moduly ke strohému

jádro dodá.

PostgreSQL

Pro perzistentní uložení informací bylo nutné zvolit některý z relačních databázových enginů. V tomto případě padla volba na PostgreSQL, což je open-source relační SQL databáze. Stejně tak by pro potřeby projektu byla vyhovující jakýkoliv Flaskem podporovaný engine, například MariaDB nebo MySQL.

Celery

Důležitým vlastností pro některé aplikace je možnost vykonávat některé operace asynchronně. K tomu lze v případě Flasku použít například Celery. Jedná se distribuovaný systém sloužící pro zpracování zasílaných zpráv, které předávány skrze nějakého z podporovaných prostředníků, tzv. brokerů. I když primárně míří na zpracování v reálném čase, podporuje například zařazení obdržených zpráv do fronty a jejich sekvenční zpracování.

Chceme-li vytvořit nějakou metodu, kterou bude možné v pythonu spustit asynchronně, je nutné obalit ji dekorátorem `@celery.task`. Tím je umožněno instanci Celery workeru vyhledat všechny tasky v dané aplikaci a uložit si je do paměti. Obdrženou zprávu lze poté chápat jako dvojici název metody-argumenty. Pracujeme-li s databázovými objekty, je možné k jejich předávání přistoupit dvěma způsoby. Buď jako argument předáme ID objektu a v metodě samotné je objekt načítán z databáze a nebo předáme jako argument objekt samotný, který je následně serializován před zasláním zprávy.

Redis

Jako broker pro komunikaci s Celery byl použit Redis. Jedná se o úložiště pracující primárně v operační paměti, nicméně lze jej nakonfigurovat tak, aby ukládal data na pevný disk a po restartu zařízení byl schopen obnovit svůj stav.

jQuery

jQuery je již několik let nejpoužívanější javascriptovou knihovnou ³ vůbec. Mezi její silné stránky patří snadná manipulace s DOM nebo tvorba AJAX requestů. Těchto vlastností využívá zejména průvodce tvorbou datasetu v administrační části aplikace.

Bootstrap

Spolu s jQuery tvoří CSS knihovna Bootstrap základ velkého množství webů. Programátor je s jejím použitím odstíněn od nutnosti stylovat prvky webové stránky a je možné se více soustředit na samotný vývoj backendu.

³<https://w3techs.com/technologies/overview/javascriptlibrary/all>

PyTorch

Open-source knihovna zaměřená na strojové učení v jazyce Python.

3.2 Prolamované typy

- **adiseet.mfcr.cz** - Daňový portál Finanční správy České republiky.
- **kamody.cz** - Typický příklad e-shopu s nevyhovujícím zabezpečením.
- **mojedatovaschranka.cz** - Portál určený pro elektronickou komunikaci s úřady.⁴
- **telerik.com** - Domovská stránka frameworku pro ASP.NET, který využívají miliony webů.
- **ulozto.cz** - Populární služba pro sdílení souborů.

3.3 Extrakce symbolů

3.3.1 Podporované grafické operace

-

3.4 Uživatelská část

Část aplikace, která je dostupná široké veřejnosti, lze rozdělit do dvou skupin – prezentace známých schémat a API. Domovská stránka seznamuje uživatele zejména s účelem projektu a dává mu možnost vyzkoušet si prolamování CAPTCHA obrázků v praxi. Po výběru obrázku k prolomení a příslušného klasifikátoru je možné zaslat dotaz na server, který z obrázku dekoduje znaky pomocí vybraného klasifikátoru, což je řešeno asynchronním dotazem na API rozhraní. Nastane-li chyba, je uživateli zobrazena příslušná hláška. V případě úspěšného rozpoznání pouze obsah zasláného obrázku.

API obsahuje pouze jednu cestu, na kterou se zasílá POST request se dvěma parametry – ID klasifikátoru a obrázek zakódovaný do Base64. Jakmile server obdrží požadavek, nejprve zkontroluje existenci klasifikátoru. Pokud přijal špatnou hodnotu, informuje o tom v odpovědi. Následně dekoduje obrázek, zjistí jeho formát (jpeg, png, gif, etc..) a normalizuje jej do podoby kterou akceptuje klasifikátor, resp. příslušný proces extrakce symbolů.

⁴V průběhu tvorby diplomové práce došlo k zásadní aktualizaci webové aplikace `mojedatovaschranka.cz` a původní CAPTCHA zabezpečení bylo nahrazeno technologií Google reCAPTCHA.

URL	metoda	parametry
/	GET	-
/decode/	POST	image, classifier ID

Tabulka 1: URL endpointy pro globální namespace

```
{
  "message":
    "Unknown classifier",
  "status":
    "error"
}
```

Chybová odpověď

```
{
  "message":
    "10487",
  "status":
    "success"
}
```

Odpověď v případě rozpoznání znaků

3.5 Administrační rozhraní

Část webu dostupná pouze administrátorovi je dostupná na URL /admin. Všechny dotazy přicházející na adresy v tomto jmenném prostoru musejí být autentizovány. Jelikož se nepředpokládá užívání aplikace více administrátory, jsou údaje poskytnuté na přihlašovací stránce kontrolovány pouze vůči hodnotám APP_USERNAME a APP_PASSWORD v konfiguračním souboru. Bezpečnost takového přístupu lze dále zvýšit například restrikcí IP adres u dotazů přicházejících na tyto adresy v konfiguraci použitého aplikačního serveru.

Nahrání datasetu /admin/datasets/new/

Proces nahrání datasetu obsahuje konfigurátor vytvořený pomocí knihoven jQuery a Bootstrap. Administrátor je nejprve vyzván k vybrání ZIP archivu se vzorovými CAPTCHA obrázky. Následně je mu umožněno označit nahrané obrázky texty, které obsahují. V posledním kroku je konfigurátor operací, díky nimž je možné zadefinovat pro nahrané schéma postup odstranění šumu pro extrahování symbolů. Dostupné jsou zejména tzv. morfologické operace, nicméně aplikace je navržena univerzálně a po vytvoření nové třídy reprezentující operaci dědicí z třídy `AbstractOperation` je možné výběr rozšířit. Jakmile je administrátor spokojený s výsledkem, odešle dataset na server a je přesměrován na stránku s jeho detaily.

Zobrazení datasetu /admin/datasets/:id/

Jakmile je dataset nahrán na server, je administrátor přesměrován na stránku s jeho detaily. V horním řádku stránky jsou informace o času vytvoření, znaky obsažené v datasetu a celkový počet rozpoznávaných symbolů. Následuje výpis nahraných obrázků spolu se zobrazením extrahovaných symbolů. V poslední části je vypsaná konfigurace extraktoru ve formátu JSON.

Tvorba klasifikátoru `/admin/classifiers/new/`

Formulář konfigurace klasifikátoru je parametrizován čtyřmi vstupy - název klasifikátoru, maximální počet iterací, cílová hodnota ztrátové funkce a zdrojový dataset. Po odeslání požadavku na vytvoření klasifikátoru je samotná fáze trénování spuštěna na pozadí a administrátor je přesměrován na nástěnku, kde vidí průběh trénování.

Detaily klasifikátoru `/admin/classifier/:id/`

Stránka zobrazující detaily klasifikátoru v současné době zobrazuje informace z fáze učení klasifikátoru. Tedy kromě názvu klasifikátoru a zvoleného datasetu jde o výslednou hodnotu ztrátové funkce a stav fáze učení klasifikátoru. Kromě toho je možné již nahraný dataset odstranit.

Nástěnka `/admin/overview/`

Jelikož se prvky nástěnky odkazují na pojmy vyjmenované výše, je popis domovské stránky administrátora až poslední. Na nástěnce jsou zobrazeny statistiky příchozích dotazů na rozpoznání CAPTCHA obrázků a počtu datasetů, resp. klasifikátorů. Existují-li navíc klasifikátory, které se momentálně trénují, jsou na nástěnce zobrazeny informace o stavu procesu. Kromě názvu klasifikátoru se jedná o pořadí současné iterace fáze učení nebo aktuální hodnotu ztrátové funkce. Tyto informace se pravidelně obnovují každých 5 sekund po asynchronním vyžádání detailů. Jakmile je klasifikátor natrénován, záznam z nástěnky zmizí.

URL	metoda	parametry
/	GET	-
/overview/	GET	-
/datasets/	GET	-
/datasets/new/	GET, POST	-
/datasets/new/preview	POST	image, operations
/datasets/:id/	GET	-
/datasets/:id/delete/	POST	-
/classifiers/	GET	-
/classifiers/new/	GET, POST	-
/classifiers/:id/	GET	-
/classifiers/:id/delete/	POST	-
/task_status/:id/	GET	-

Tabulka 2: URL endpointy pro namespace `/admin/`

Další vývoj

Jak je patrné již z popisu aplikace, k vytvoření ultimativního nástroje pro lámání CAPTCHA kódů, byť pouze obrázkových, vede ještě dlouhá cesta. Implementace následujících funkcionalit, zejména v administračním rozhraní, by výrazně zlepšila komfort užívání a zřejmě i praktičnost systému jako celku.

- **Upozornění na chybně klasifikované znaky.** Na stránce zobrazující detaily klasifikátoru se naskýtá možnost zobrazit chybně klasifikované obrázky. Taková chyba může nastat zejména v následujících třech případech: 1) Příliš podobné znaky (číslíce nula versus písmeno O). 2) Algoritmus pro extrakci symbolů na výstupu nevrátí celý znak, ale např. pouze jeho část. 3) Symbol byl ručně chybně označen a klasifikátor jej rozpoznává správně.
- **Změna označení již nahraných obrázků.** Během procesu vytváření datasetu musí administrátor nahrát ZIP archiv obsahující obrázky jejichž název odpovídá CAPTCHA kódu, nebo je procesem označení obrázků proveden před uploadem na server. V obou případech však závisí jen a pouze na lidském faktoru, zda bude obrázku přiřazen chybný text. To je například poslední bod předchozího odstavce. Z toho důvodu je vhodné umožnit změnu označení obrázku, resp. samotného symbolu.
- **Odstranění některých znaků.** Druhá chyba z prvního odstavce popisuje nesprávné rozpoznání symbolu v CAPTCHA obrázku. Tomu je možné předejít úpravou parametrů extraktoru, což však může mít za následek horší rozpoznávací schopnosti symbolů v datasetu jako celku. Abychom zamezili trénování klasifikátoru na chybných datech, je vhodnější poskytnout možnost odebrat jednotlivé symboly z datasetu, které byly špatně rozpoznány. Příímý důsledek tohoto řešení je nutná úprava procesu trénování klasifikátoru CNN, resp. hodnoty udávající počet současně zpracovávaných vstupů.
- **Volba z více klasifikátorů.** V současné době systém používá jako klasifikátor konvoluční neuronovou síť se dvěma konvolučními vrstvami, kde se na každý vstup uplatní funkce `MaxPool` s velikostí jádra = 2. To nemusí vyhovovat všem rozpoznávaným instancím CAPTCHA obrázků a i pro účely porovnání efektivnosti klasifikátorů je rozumné přidat další a mít možnost volby.
- **Volba tvaru jádra morfologických operací.** Každá morfologická operace potřebuje dva vstupy - obrázek a jádro. Tvar, resp. hodnota matice, jádra má přímý vliv na výsledný obrázek. Aplikace parametrizuje jádro pouze jednou celočíselnou hodnotou udávající velikost čtvercové jednotkové matice. Je však možné použít jádro ve tvaru tzv. kříže, resp. elipsy. Příklad lze vidět na obrázku níže.

- **Klasifikátor používající více datasetů.** K dalšímu zlepšení rozpoznávání by mohlo také vést trénování klasifikátoru znaky z více zdrojů. Kromě primárního datasetu, který udává i proces extrakce symbolů, by k jeho množině znaků byly přidány ty symboly s odpovídající hodnotou ze zvolených datasetů. Tím by se docílilo rozpoznání transformací, které v trénovací množině symbolů nebyly prve zahrnuty.

Otázky

- n-fold u klasifikátoru? s tím souvisí úspěšnost klasifikátoru
- experimentální porovnání – nn vs cnn, rozpoznání na úrovni obrázků vs znaků
- počet stran práce
- u datasetu zobrazit pokrytí jednotlivých symbolů
- paginace obrázků datasetu

Závěr

Závěr práce v „českém“ jazyce.

Conclusions

Thesis conclusions in “English”.

A První příloha

Text první přílohy

B Druhá příloha

Text druhé přílohy

C Obsah přiloženého CD/DVD

Na samotném konci textu práce je uveden stručný popis obsahu přiloženého CD/DVD, tj. jeho závazné adresářové struktury, důležitých souborů apod.

bin/

Instalátor `INSTALATOR` programu, popř. program `PROGRAM`, spustitelné přímo z CD/DVD. / Kompletní adresářová struktura webové aplikace `WEBOVKA` (v ZIP archivu) pro zkopírování na webový server. Adresář obsahuje i všechny runtime knihovny a další soubory potřebné pro bezproblémový běh instalátoru a programu z CD/DVD / pro bezproblémový provoz webové aplikace na webovém serveru.

doc/

Text práce ve formátu PDF, vytvořený s použitím závazného stylu KI PřF UP v Olomouci pro závěrečné práce, včetně všech příloh, a všechny soubory potřebné pro bezproblémové vygenerování PDF dokumentu textu (v ZIP archivu), tj. zdrojový text textu, vložené obrázky, apod.

src/

Kompletní zdrojové texty programu `PROGRAM` / webové aplikace `WEBOVKA` se všemi potřebnými (příp. převzatými) zdrojovými texty, knihovnami a dalšími soubory potřebnými pro bezproblémové vytvoření spustitelných verzí programu / adresářové struktury pro zkopírování na webový server.

readme.txt

Instrukce pro instalaci a spuštění programu `PROGRAM`, včetně všech požadavků pro jeho bezproblémový provoz. / Instrukce pro nasazení webové aplikace `WEBOVKA` na webový server, včetně všech požadavků pro její bezproblémový provoz, a webová adresa, na které je aplikace nasazena pro účel testování při tvorbě posudků práce a pro účel obhajoby práce.

Navíc CD/DVD obsahuje:

data/

Ukázková a testovací data použitá v práci a pro potřeby testování práce při tvorbě posudků a obhajoby práce.

install/

Instalátory aplikací, runtime knihoven a jiných souborů potřebných pro provoz programu PROGRAM / webové aplikace WEBOVKA, které nejsou standardní součástí operačního systému určeného pro běh programu / provoz webové aplikace.

literature/

Vybrané položky bibliografie, příp. jiná užitečná literatura vztahující se k práci.

U veškerých cizích převzatých materiálů obsažených na CD/DVD jejich zahrnutí dovolují podmínky pro jejich šíření nebo přiložený souhlas držitele copyrightu. Pro všechny použité (a citované) materiály, u kterých toto není splněno a nejsou tak obsaženy na CD/DVD, je uveden jejich zdroj (např. webová adresa) v bibliografii nebo textu práce nebo v souboru `readme.txt`.

