

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

Tím č.5
Zápisnica č.10

Dátum a čas: 23.11.2020, 16:00 - 19:00

Vedúci tímu: Mgr. Martin Sabo, PhD., Ing. Marta Prnová, PhD.

Kontakt: mikasa.fiit@gmail.com

Vypracovala: Mária Rajníková

1 Zhrnutie stretnutia

Prítomní na stretnutí

Vedúci tímu:

- Mgr. Martin Sabo, PhD.
- Ing. Marta Prnová, PhD.

Členovia tímu:

- Simona Klučková
- Maroš Kollár
- Jakub Kučečka
- Zuzana Popovcová
- Mária Rajníková
- Alena Valová

Hlavné body stretnutia

1. Sprint review
2. Schvaľovanie user stories 4 šprintu
3. Mock-ups
4. Riešenie otázok členov tímu

2 Priebeh stretnutia

Sprint review

Na začiatku stretnutia sme prezentovali výsledky tretieho šprintu vedúcim tímového projektu. Z klasifikácie sme sa predstavili ako prvé nové črty a ich úspešnosti:

- min, median, std, delta, sum
- point
 - priemer/maximum s vybranej oblasti
 - úspešnosť: 52-78
- diff
 - o rozdiel od šablóny (priemer máp danej triedy oleja)
 - o úspešnosť: 56-73

Pri hľadaní anomálií pre črtu point sme orezali celkové dáta na niekoľko malých vybraných oblastí. Tieto oblasti predstavujú veľmi malú časť, napriek tomu majú porovnateľnú úspešnosť s doterajšími výsledkami, vďaka čomu môžeme začať so znižovaním objemu dát na spracovanie.

Ako ďalšia bola predstavená konvolučná neurónová sieť. Pri jej implementácii sa vyskytol problém s veľkosťou pamäte a riešenie so znižovaním obrázkov. Celkovo máme úspešnosť 20-74%, no najväčším problémom je predikcia zameraná na jednu triedu. Riešili sme otázku, či pokračovať a dohodli sme sa ďalej skúšať a vedúci nám odporučili skúsiť prístupniť školský počítač.

Dnes odovzdávame dokumenty po treťom šprinte. Predstavili sme ich vedúcim a upozornili na termín odovzdania.

Schvaľovanie user stories 4 šprintu

Prezentovali sme user stories a tasky na nasledujúci šprint. Všetky boli schválené vedúcimi.

1. Klasifikácia

- (a) Úprava vstupných súborov
- (b) Vylepšenie modelov
- (c) Refaktorizácia klasifikácie

2. Aplikácia

- (a) Obrazovky

3. Administrácia

- (a) Kompletizácia odovzdania v 9-tom týždni
- (b) Zabezpečenie prípravy dokumentácie a stretnutí

User stories boli schválené a prešli sme k otázkam. Martin Sabo sa spýtal, či je potrebné dokúpiť RAM, načo Maroš Kollár odpovedal, že to pravdepodobne nepomôže, kvôli predikcií jednej triedy. Skúsime ešte, aké budú výsledky po odstránení stĺpcov v ďalšom šprinte a pre iné druhy sietí. Marta Prnová navrhla pozrieť sa na riešenia cez histológiu.

Mock-ups

Prezentovali sme návrh obrazovky mock-up vedúcim. Martin Sabo mal pripomienku ohľadom zobrazenia obrázkov na šírku, nakoľko uprednostňuje na výšku a taktiež či ich chceme zobrazovať.

Ďalej pokračuje voľnejšia debata návrhu jednotlivých obrazoviek. Obrazovky sa viackrát prerábali, hlavné body a ich riešenia sú nasledovné:

1. Spúšťa sa klasifikácia vždy nad všetkými dátami?
 - iba tie čo označí
2. Kedy sa zobrazujú obrázky
 - Obrázky sa zobrazujú už po spracovaní
3. Čo zobraziť na detaily vzorky?

- ktoré klasifikátory boli použité
- percentuálna šanca pri Random Forest
- odkaz na zobrazenie obrázka
- do budúcnosti možno vyznačenie číť, ktoré sa použili

4. Chceme nahratie modelov?

- odstrániť nahranie modelu nakoľko nie je vhodné nahrávať si vlastné ťažko overiteľné modely, keď ponúkame vlastné

Ďalej definované riešenia

- Nebudeme generovať obrázky pre všetky oleje naraz
- Trénovať nový - natrénuje vybraný model na nových dátach a uloží ako nový
- Dotrénovať model – dotrénuje vybraný model a uloží ako nový
- Pridať zoradenie modelov od najnovších a čas vytvorenia
- Nebudeme meniť existujúce modely
- Umožniť vymazávanie modelov, pričom nie je možné vymazať základné modely (aby mal používateľ s istotou funkčné)
- Jazyk aplikácie bude slovenčina a angličtina. Výber bude realizovaný ako dropdown.
- Hlavný panel sa presunie hore, štýl ikonka s popisom
- Vytváranie pokračovalo ako posledný bod stretnutia, kde sme spoločne vytvorili novú verziu obrazoviek.

Riešenie otázok členov tímu

Front-end

Jakub Kučička predstavil problémy ako spraviť front-end pre Python a niekoľko riešení. Možnosti sú čistý tkinter a HTML.

Dohodli sme sa že je lepšie HTML, aj keď má niekoľko komplikácií, ktorým sa budeme venovať ešte v tomto šprinte, pozrieme sa na existujúce možnosti a skúsime vytvoriť exe súbor a posielanie podľa json.

Jeden z návrhov bol zmeniť jazyk pre front-end, ktorý sa zamietol pre kompatibilitu.

Konfiguračný súbor

V tomto šprinte sa vytvára konfiguračný súbor, ktorý obsahuje globálne premenné. Prvou otázkou je ako rozdeliť tagy pre premenné. Dohodli sme sa na raw, original, features pre prístup k dátam, pričom podľa počtu premenných sa môžu ešte upraviť. Pre ostatné premenné zostáva rozdelenie rovnaké ako pre priečinky: Preprocessing, Features...

Ďalej sa rozoberalo, či volať config na naplnenie globálnej premennej alebo priamo na mieste, kde potrebujeme hodnotu a schválila sa druhá možnosť. Jakub Kučička mal poznámku či by sme nevolali config len raz a nie v každom súbore. Vzhľadom na to, že funkcie ešte nie sú plne prepojené a nie je možné ho poselať nechávame volanie configu v každom súbore a neskôr môžeme skúsiť, či sa to bude dať inak.

Diagramy funkcií

Zuzana Popovcová prezentovala existujúce diagramy, ktoré sme zmenšovali, aby neboli na celú stranu. Diagramy sú rozdelené podľa sekcií. Mária Rajníková sa pýtala, či nedať obrázky medzi opísané funkcie v dokumentácií. Dohodli sme sa, že sa nič nemení nakoľko obrázky do popisu funkcie nepatria.

Nový súborový systém

Mária Rajníková opísala nové rozdelenie dát po rozdelení na tréningové a testovacie vzorky. Opisovali sme veľkosti súborov, s ktorými budeme odteraz robiť.

Dohodli sme sa, že hyperparametre sa budú tréňovať až po výbere čít (features selection).