

Slovenská technická univerzita v Bratislave  
Fakulta informatiky a informačných technológií

Tím č.5  
Zápisnica č.9

Dátum a čas: 16.11.2020, 16:00 - 19:00

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

Kontakt: mikasa.fiit@gmail.com

Vypracoval: Jakub Kučečka

# 1 Zhrnutie stretnutia

---

## Prítomní na stretnutí

Členovia tímu:

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

## Hlavné body stretnutia

1. Poznámky k tvorbe aplikácie
2. Problémy s pamäťou
3. Diskusia ku klasifikácii
4. Inžinierske dielo

## 2 Priebeh stretnutia

---

Pred koncom šprintu sme narazili na viacero otázok, ktoré nám blokovali zatváranie taskov. Preto bolo toto stretnutie zamerané prevažne na riešenie týchto otázok.

### Poznámky k tvorbe aplikácie

Začali sme návrh aplikácie, kde sme zistili, že nemáme kompletne vyriešené všetky náležitosti. Dohodli sme sa teda, že dokument Riadenia projektu bude obsahovať funkcionálne a nefunkcionálne požiadavky, ktoré sú už vo finálnej podobe a nie je potrebné ich upravovať. Dohodli sme sa, že Zuzana Popovcová ešte dokončí akceptačné testy, na koľko sme na zopár maličkostí nemysleli.

So spísanými prípadmi použitia sme mali veľa problémov. Dohodli sme sa teda, že Jakub Kučečka zapracuje tieto zmeny:

- Zahrnie výber viacero súborov.
- Zapracuje validáciu .txt súborov.
- Opraví alternatívny scenár pre UC1.
- Opraví UC na Prípad použitia.
- Pridá alternatívny scenár do UC2.
- Odstráni prípony .png a .csv.
- Zapracuje tlačidlo optimalizácia hyperparametrov.
- Z UC4 a UC2 vyberie spoločné kroky.

Dohodli sme sa, že Mária Rajníková v spolupráci so Zuzanou Popovcovou vytvorí diagramy interakcií funkcií.

## Problémy s pamäťou

Pokračovali sme s implementáciou klasifikácie. Tu sme zistili, že pravdepodobne zahrála do karát zhoda okolností spojená s týmto šprintom. Čo sa teda mohlo pokaziť tak sa pokazilo. Vyhodnotili sme, že piatok trinásteho sa asi naozaj nemá pracovať.

Prvým problémom bola normalizácia, kde sme znásobili objem súborov, nakoľko pôvodné hodnoty mali od 3 do 6 desatinných miest a výsledne mali 9 desatinných miest. Tu sme ale vyhodnotili, že sa na tom spálilo dostatok času a po orezaní dát nebudú nároky na pamäť také tragické. Diskutovali sme ešte o zaokrúhlení hodnôt, čo sme ihneď zamietli, z dôvodu straty údajov.

Ďalším problémom bola neurónová sieť VGG16. Nakoľko pôvodné obrázky mali rozmer 9000x449 neurónová sieť nedokázala vziať do pamäte ani 4 obrázky naraz. Pri 3 obrázkoch zasa odmietala prechod cez 13. vrstvu. Výsledkom, ktoré je zatiaľ akceptované je, že obrázky sme zmenšili na rozmer 224x224 čím sme pravdepodobne stratili veľa údajov. Rozhodli sme sa ešte skúšať väčšie obrázky dokedy nám to povolí pamäť, ale toto nebude prebiehať v tomto šprinte. Zahrnuli sme aj testovanie siete týmito obrázkami a výsledok je taký, že do 6 epochy sieť nedosahovala chvályhodné výsledky a od 8 epochy sa sieť začínala pretrénovávať, preto sme predbežne vyhodnotili, že konvolučné siete nebudú správna cesta.

Ďalšie problémy s pamäťou nastali na pri tasku s vytváraním črt. Nakoľko pri 5. črtách nad 447 stĺpcami týkajúcich sa triedy EVOO pracujeme nepretržite s pamäťou vyše 15GB. Preto sa nám ich nepodarilo všetky vytvoriť.

Navrhlo sa teda viacero riešení ako prácu s pamäťou zjednodušiť. Navrhlo sa:

- vyhodiť t-stĺpce
- rozdelenie na test a train adresáre hneď na začiatku
- orezať vstupné dáta hneď na začiatku
- viac RAM

Pri týchto návrhoch sme sa zatiaľ rozhodli, že zapracujeme prvé dva body.

## **Diskusia ku klasifikácii**

V tejto diskusii sa zistilo, že chunkovanie súborov, je zatiaľ nereálne, nakoľko na predspracovanie dát a normalizáciu nemôžeme postupovať ani po riadkoch a ani po stĺpcoch daných tried olejov. Dohodlo sa teda, že Simona Klučková spíše prečo sa to robiť nedá do tasku.

Pri prezentácii diff máp sme zistili, že toto je krok správnym smerom, nakoľko modely natréované iba na týchto črtách dosahujú chvályhodné výsledky. Preto sme sa rozhodli ďalej pracovať na týchto črtách. Tiež sme zistili, že podstatnú časť dát nepotrebujeme, pretože sú pre všetky triedy rovnaké.

## **Inžinierske dielo**

Pri tvorbe dokumentu vznikli problémy s neefektívnymi WBS diagramami. Preto sme sa rozhodli pre ich modifikáciu, ktorú zapracuje Alena Valová. Tiež sme sa dohodli, že metodiky pridá do príloh a odkážeme sa na nich v texte, zároveň odstráni farebné vyznačenie odkazov a aj samotné odkazy pre softvéry tretích strán.