

ИЗВЕШТАЈ

о прегледу мастер рада

„Имплементација и евалуација алгоритама машинског учења за филтрирање нежељене поште“

кандидата Александра Недељковића

Одлуком Наставно-научног већа Математичког факултета која је донета на 322. редовној седници одржаној 19. јуна 2015. године именовани смо за чланове комисије за преглед и оцену мастер рада под насловом „Имплементација и евалуација алгоритама машинског учења за филтрирање нежељене поште“ кандидата Александра Недељковића, студента мастер студија на студијском програму Информатика на Математичком факултету.

I Област рукописа

У рукопису „Имплементација и евалуација алгоритама машинског учења за филтрирање нежељене поште“ кандидат Александар Недељковић бави се описом техника машинског учења и алгоритама класификације заснованих на Бајесовом филтрирању, методи К-најближих суседа, методи подржавајућих вектора и вештачким неуронским мрежама и применом тих алгоритама на проблем филтрирања нежељене поште. У раду се користи знање машинског учења, алгоритама и програмирања.

II Структура рукописа и кратак приказ

Рукопис се састоји од 85 страна, организованих у 7 поглавља, библиографију и додатак.

У уводном поглављу кандидат описује појам комуникације електронском поштом и проблем нежељене поште .

У поглављу „Класификација“ кандидат дефинише појам класификације текстуалних докумената и мере квалитета и технике евалуације класификатора.

Поглавље „Алгоритми филтрирања електронске поште“ је једно од централних поглавља у раду и у њему кандидат детаљно описује технику Бајесовог филтрирања, методу К-најближих суседа, методу подржавајућих вектора и вештачке неуронске мреже.

У поглављу „Подаци“ кандидат описује јавно доступне корпусе порука који су коришћени за евалуацију техника филтрирања (корпусе PUA1, PUA2, PUA3, PUA, Enron-Spam и Ling-Spam).

У поглављу „Имплементација“ кандидат описује своју имплементацију алгоритама филтрирања нежељене поште. У оквиру имплементације коришћена је библиотека SVM^{light} која је укратко и описана.

У поглављу „Експериментални резултати“ кандидат даје табеларни и графички приказ резултата тестирања своје имплементације алгоритама филтрирања поште на корпусима описаним у поглављу „Подаци“.

У поглављу „Закључци“ кандидат изводи закључе на основу резултата изведених експеримената. Разни приступи се пореде на основу времена потребног за тренирање, као и добијене тачности истренираног класификатора.

Библиографија садржи списак од 32 библиографске јединице које је кандидат користио приликом писања рада.

Додатак садржи приказ делова изворног кода програма.

III Анализа рукописа

У рукопису који смо анализирали кандидат је описао неколико техника машинског учења које се обично користе у имплементацији програма за детекцију и класификацију нежељене електронске поште. На основу прикупљених теоријских описа техника имплементиран је програмски систем који може да примени сваку од описаних техника на дати корпус порука електронске поште (систем подржава фазу тренинга, а затим и експлоатацију истренираних класификатора, коа и класификацију комбиновањем више метода). Због комплексности имплементације методе подржавајућих вектора, коришћена је јавно доступна библиотека SVM^{light}. Имплементација је прилично опсежно тестирана на неколико јавно доступних корпуса електронске поште (укупан број порука у свим корпусима је око 40.000). Прецизност и одзив се разликују од корпуса до корпуса, као и од методе до методе, што је и очекивано. Главни закључак који кандидат изводи је да се највећа тачност класификације и најмање време тренирања постиже код методе подржавајућих вектора и код вештачких неуронских мрежа (перцептрона) и препоручује њихову употребу за решавање проблема филтрирања нежељене поште.

IV Закључак и предлог

Описом, имплементацијом и детаљном евалуацијом различитих техника филтрирања нежељене поште заснованих на машинском учењу кандидат је показао задовољавајући степен стручног знања. На основу наведеног Комисија предлаже да се рукопис под насловом:

„Имплементација и евалуација алгоритама машинског учења за филтрирање нежељене поште“

кандидата Александра Недељковића прихвати као мастер рад и да се закаже његова јавна одбрана.

Комисија:
др Филип Марић, ментор

др Младен Николић

др Јелена Граовац

Београд, 27. 1. 2016.