

POSUDEK VEDOUCÍHO BAKALÁŘSKÉ PRÁCE

Autor práce: Petr Kubelka

Název práce: Semantic Biclustering Optimization

Vedoucí: doc. Ing. Jiří Kléma, Ph.D., katedra počítačů, FEL ČVUT

Hlavním cílem bakalářské práce pana Kubelky byl návrh, implementace a experimentální ověření nové varianty algoritmu sémantického dvojshlukování. Ve stručnosti jde o současné shlukování podobných pozorování ve dvou dimenzích. To není řízeno pouze podobností samotných pozorování, ale i sémantickou podobností popisů v obou dimenzích. Úlohu student řešil v konkrétní doméně molekulární biologie, kde pozorováními jsou hodnoty genové exprese měřené v různých biologických situacích (dimenze 1) u různých genů (dimenze 2). Cílem je nalézt dvojshluky genů a situací pro které platí, že situace i geny mají homogenní popis a exprese se výrazně neliší.

Student vycházel ze stávající implementace algoritmu, kdy se kritéria sémantické homogenity a homogenity exprese uplatňovala postupně. Nejprve se našly dvojshluky s homogenní expresí, poté algoritmus vyhledával jejich zobecnění s uniformním popisem. Úkolem studenta bylo k problému přistoupit jako k multikriteriální optimalizační úloze a od úvodních fází algoritmu zohledňovat všechna kritéria. Pan Kubelka neřešil samotnou optimalizační úlohu, zaměřil se na problém agregace alternativních řešení této úlohy na malou množinu nepřekrývajících se dvojshluků poskytujících biologicky relevantní informaci.

Pan Kubelka řešil bakalářskou práci dlouhodobě a systematicky, kvalitu naší spolupráce hodnotím kladně. Všechny základní cíle bakalářská práce splnila. Student se seznámil s problematikou v celé své šíři (shlukování a jeho rozšíření, práce s anotacemi ve formě ontologií, optimalizace, ověření výsledků pomocí klasifikace), algoritmus naimplementoval a nezanedbatelné úsilí věnoval i jeho netriviálnímu ověření. To samo o sobě při složitosti úlohy zaslouží ocenění. Na druhou stranu si uvědomuji, že pan Kubelka byl téměř výhradně implementátorem a experimentátorem. Pokud jde o algoritmus, zůstali jsme u počátečního návrhu a nepodařilo se jej na základě průběžných výsledků nijak rozvinout. Současné výsledky přitom naznačují, že další úpravy algoritmu budou nutné. Bylo tomu i proto, že zásadní pokrok zaznamenal pan Kubelka zejména v posledních 3 měsících řešení a na kritickou revizi algoritmu nezbyl čas. Pokud jde o text práce, z počátku málo uspokojivý text se autorovi v posledních týdnech podařilo vylepšit do rozumné formy.

Bakalářská práce splňuje všechny body zadání a má podle mého názoru dobrou úroveň. Student prokázal schopnost samostatné práce. Ze všech výše uvedených důvodů práci hodnotím známkou

C — dobře.

V Praze, dne 4. června 2018

doc. Ing. Jiří Kléma, Ph.D., vedoucí