

## Posudek na diplomovou práci Bc. Štefana Okruhlici

Práce se zabývá různými přístupy ke clusterování a následné vizualizaci nominálních a ordinálních dat. Jedná se o poměrně komplexní problém a je třeba hned v úvodu posudku říci, že se jeho zpracování v práci je na velmi dobré úrovni.

Práce má dobrou strukturu, kde jsou nejprve popsány různé vizualizační metody a z nich diplomant vybral ty, které pokládá za vhodné k implementaci zvoleného řešení. V úvodu práce je provedena analýza charakteristik nominálních a ordinálních dat, která budou vizualizována. Výklad je doprovázen řadou názorných obrázků dokumentujících vlastnosti jak vizualizovaných dat, tak i vlastnosti jednotlivých vizualizačních metod.

Za obzvláště přínosné pokládám aplikaci metod *hierarchical clustering* a *paralelogram clustering* spolu s automatickým uspořádáním atributů a jejich kategorií. Diplomant testoval různé přístupy, přičemž výsledkem testů byla volba optimální metody použité v implementaci. Testování probíhalo na datovém souboru o pasažérech Titanicu. Výsledky testů byly dobře okomentovány.

Za obzvláště cennou pak pokládám kapitolu 4 (Experimental results), kde je implementovaná metoda ověřována na dalších datových souborech. Výsledky testů potvrdily správnost zvoleného přístupu. Práce je napsána v dobré angličtině a má velmi dobrou grafickou úpravou. Práce je napsána přehledným a čitelným způsobem (což čtenář jistě uvítá – vzhledem k poměrné složitosti tématu).

Přes všechnu chválu mám k práci několik připomínek. Navržená a implementovaná metoda dává uživatelům poměrně mocný nástroj pro vizualizaci specifických dat. Diplomant při vyhodnocování výsledků testů oceňuje různé varianty poněkud osobně zabarvenými slovy – viz str. 62: I consider threshold 0.3 the best. Chybí tak jistá metrika, podle které by se kvalita vizualizace objektivněji posuzovala. Nicméně jest pravdou, že by tak došlo k dalšímu nárůstu již tak rozsáhlé diplomové práce (vypracování metriky a její otestování by určitě vyžadovalo jisté úsilí). Obdobný problém je i na str. 31: silhouette is 0.7 which is quite good (od které meze je hodnota nedobrá?).

V závěrečné kapitole (Conclusion) jsou hezky naznačeny směry dalšího rozvoje vyvinuté metody. Poněkud mi chybí následující úvaha: vizualizace slouží uživatelům s cílem lepší orientace v datech – implementovaný systém by měl projít i testováním ze strany uživatelů, aby se zjistilo nakolik se uživatelé lépe orientují v datech (tato poznámka má souvislost s mou předchozí poznámkou).

Kromě výše uvedených připomínek mám i několik připomínek formálního rázu.

V rozsáhlejších textech typu diplomová práce apod. rozděleným na kapitoly je dobrým zvykem, že jednotlivé kapitoly mají, pokud možno, srovnatelnou délku. Zde tomu tak není. V kapitole 3 popisující implementaci je část, kde se popisují datové typy a jiné implementační podrobnosti. Bývá dobrým zvykem přesunout informace tohoto typu do přílohy (tělo práce popisuje spíše koncepční věci). Při popisu uživatelského rozhraní jsou popisovány jen rozšiřující funkce bez toho, aby byly popsány i funkce původní (i tato část měla být dána do přílohy). Na str. 53 jsou popisovány další datové soubory, se kterými bude provedeno ověřování funkčnosti implementace. Chybí zmínka o tom, kde se tyto soubory vzaly a jejich alespoň základní charakteristika (počet položek atd.).

Diplomová práce má nadstandardní úroveň. Diplomant pracoval dobře s literaturou. Přes několik mých připomínek považuji práci za vynikající, a proto navrhuji hodnocení A (výborně).

Praha 21.8.21

prof. ing. Pavel Slavík, CSc.

