

Posudek vedoucího diplomové práce

Téma: Sampling-based Motion Planning for Tunnel Detection in Protein Structures

Student: Jindřich Durčák

Posudek vypracoval: Ing. Vojtěch Vonásek, Ph.D., Katedra kybernetiky, ČVUT FEL

Diplomová práce se zabývá dvěma hlavními tématy: výpočtem tunelů v proteinových strukturách a testováním průchodnosti ligandů tunelem. Jedná se o velmi aktuální problémy, jejichž řešení může pomoci v celé řadě oblastí, např. při výzkumu léků. Cílem této práce bylo vyzkoušet detekci tunelů s využitím algoritmů známých z robotického plánování pohybu. Student se zaměřil na tzv. sampling-based metody.

Přehled práce je uveden v první kapitole. V druhé kapitole je čtenář velmi pěkně seznámen se základy biochemie. Zde student zhodnotil znalosti z předchozího studia. Této části si velmi vážím, neboť poslouží jako dobrý startovní bod pro další studenty. State-of-the-art metody pro výpočet tunelů jsou uvedeny ve třetí kapitole. Popis state-of-the-art metod je přehledný a nechybí zde nejvýznamnější práce.

Navržené algoritmy jsou popsány ve čtvrté kapitole. Student navrhuje několik modifikací RRT pro detekci tunelů. Navržené metody jsou podrobně popsány, uvedeny též ve formě pseudokódu a doplněny vhodnými obrázky. Jednou z důležitých modifikací je detekce povrchu proteinu, k čemuž je využit výpočet obálky proteinu metodou α -shape. Pro urychlení následných výpočtů byl navržen Algoritmus 2 (str. 45), který filtruje body originálního α -shape. Myslím, že bylo dobré doplnit jej obrázkem s ukázkou původních dat (tj. plného α -shape) a dat po filtraci, případně ještě uvést množství odfiltrovaných bodů pro proteiny použité v experimentech.

Dále byla navržena a implementována metoda pro testování průchodnosti ligandu tunelem. Zde je využito principu tzv. guided sampling, kdy jsou náhodné body generovány s větší pravděpodobností v koulích reprezentujících tunel. Tato modifikace (Algoritmus 6, str. 52) je sice názorně popsána, ale čtenáři nemusí být jasné **proč je lepší generovat náhodné body (konfigurace) v tunelu, než rovnoměrně v celém prostoru proteinu?** Algoritmus 6 generuje vzorky postupně v každé kouli reprezentující tunel. **Proč nejsou náhodné vzorky generovány v celém tunelu najednou? Jak by se to projevilo na růstu RRT stromu?** Bylo by vhodné dát na konci čtvrté kapitoly krátkou sekci se shrnutím navržených metod a jejich rozdílů oproti klasickému RRT.

Experimentální výsledky a diskuze jsou uvedeny v páté, resp. šesté kapitole. Detekce tunelů byla porovnána s nástrojem CAVER 3.0, což je jeden z nejpoužívanějších nástrojů pro detekci tunelů. Experimenty jsou podrobně popsány a výsledky detailně okomentovány. Diskuze nad dosaženými výsledky ukazuje, že student velmi dobře porozuměl jak problematice plánování pohybu, tak i jejímu použití pro detekci tunelů. Navržený algoritmus byl schopen najít nejvýznamnější tunely, což je velmi dobrý výsledek. Případy, kdy navržená modifikace RRT nenalezla žádané tunely, jsou podrobně rozebrány a jsou uvedeny možné příčiny.

V sekci 5.4 je uvedena tabulky s časovými nároky navržené metody, bohužel zde chybí časy pro CAVER1 3.0. **Kolik času potřebuje CAVER 3.0 pro detekci statických tunelů? Zahrnují časy v Tabulce 5.7 i předzpracování dat (např. výpočet α -shape a jeho filtrování)?** Uvedené časy, bez ohledu na to jestli zahrnují předzpracování dat, jsou velmi malé a prakticky zanedbatelné k celkové době výpočtu molekulární dynamiky (tam se jedná řádově o týdny).

Téma DP považuji za velmi náročné, neboť vyžaduje na jednu stranu velmi dobrou znalost robotických algoritmů plánování pohybu a schopnost navrhnout jejich modifikaci, na druhé straně také orientaci v problematice detekce tunelů, schopnost pracovat s proteinovými daty a nástroji výpočetní biochemie. Kromě implementace navržených algoritmů v C++ musel student řešit řadu technických detailů (např. příprava dat, programování skriptů pro vizualizaci výsledků v systému PyMOL), dále bylo třeba nastudovat a použít řadu knihoven, což jistě vyžadovalo mnoho času.

Spolupráce s Jindřichem Durčákem na tématu DP byla vynikající. Na konzultace chodil vždy perfektně připraven, téma zpracovával samostatně. Díky svým znalostem biochemie směřoval práci k použitelným výsledkům a kriticky se zamýšlel — a často zavrhnul — moje nápady na další zpracování tunelů, např. pokud by to z pohledu biochemie vyžadovalo přílišné zjednodušení. Jeho práce není pouhou modifikací

sampling-based metod pro doménu proteinů, lze ji použít jako dobrý základ pro další výzkumné práce. Tomu napomáhá i přehledná implementace v Python/C++ na přiloženém disku.

Práce je psána anglicky na velmi vysoké úrovni, našel jsem pouze pár drobných chybiček (např. občas je použit formát čísel 5000 a jindy 5 000, ale to jsou nepodstatné detaily). Text je logicky členěn do sekcí a vše potřebné je vysvětleno. Popisy algoritmů jsou sice stručnější, ale navržené metody podle nich naimplementovat lze. Taktéž grafické zpracování je velmi dobré. Zadané cíle práce byly splněny a i přes uvedené drobné výtky považuji práci za velmi povedenou a přínosnou.

Předloženou diplomovou práci hodnotím **A** — **výborně**.

1.6.2017, Praha

Vojtěch Vonásek