COMENIUS UNIVERSITY IN BRATISLAVA FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

EFFICIENT AND EFFECTIVE DYNAMIC TIME WARPING MASTERS THESIS

COMENIUS UNIVERSITY IN BRATISLAVA FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

EFFICIENT AND EFFECTIVE DYNAMIC TIME WARPING

MASTERS THESIS

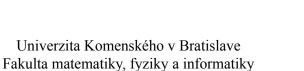
Study Programme: Computer Science Field of Study: Computer Science

Department: FMFI.KAI - Department of Applied Informatics

Supervisor: doc. Mgr. Tomáš Vinař, PhD.

Bratislava, 2020

Emanuel Tesař, Bc.





ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Emanuel Tesař

Študijný program: informatika (Jednoodborové štúdium, magisterský II. st.,

denná forma)

Študijný odbor:informatikaTyp záverečnej práce:diplomováJazyk záverečnej práce:anglickýSekundárny jazyk:slovenský

Názov: Efficient and Effective Dynamic Time Warping

Zefektívnenie algoritmu Dynamic Time Warping

Anotácia: Dynamic time warping (DTW) je technika prebratá zo spracovania reči

a často sa používa aj na zarovnávanie nanopórového signálu ku sekvenciám. Štandardná formulácia problém však nepokrýva niektoré dôležité vlastností reálnych dát, vďaka čomu výsledné zarovnania obsahujú rôzne artefakty takejto formulácie. Cieľom práce je preskúmať rôzne reformulácie problému a rozšíriť základný algoritmus tak, aby bolo možné vytvoriť praktický nástroj

na zarovnávanie nanopórového signálu ku sekvenciám.

Vedúci: doc. Mgr. Tomáš Vinař, PhD.

Katedra: FMFI.KAI - Katedra aplikovanej informatiky

Vedúci katedry: prof. Ing. Igor Farkaš, Dr.

Dátum zadania: 30.10.2019

Dátum schválenia: 21.11.2019 prof. RNDr. Rastislav Kráľovič, PhD.

garant študijného programu

študent	vedúci práce





Comenius University in Bratislava Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname: Bc. Emanuel Tesař

Study programme: Computer Science (Single degree study, master II. deg., full

time form)

Field of Study: Computer Science Type of Thesis: Diploma Thesis

Language of Thesis: English **Secondary language:** Slovak

Title: Efficient and Effective Dynamic Time Warping

Annotation: Dynamic time warping (DTW), a method from speech processing, is often used

to align nanopore signals to sequences. However, the standard formulation of the problem does not capture properly important properties of underlying data, which results in artifacts in alignments. The goal of this thesis is to study various reformulations of the problem and extend the basic algorithm to a practical tool

for aligning nanopore signals to sequences.

Supervisor: doc. Mgr. Tomáš Vinař, PhD.

Department: FMFI.KAI - Department of Applied Informatics

Head of prof. Ing. Igor Farkaš, Dr.

department:

Assigned: 30.10.2019

Approved: 21.11.2019 prof. RNDr. Rastislav Kráľovič, PhD.

Guarantor of Study Programme

Student	Supervisor

Abstrakt

DTW (z angl. dynamic time warping) je jednoduchý algoritmus na zarovnávanie dvoch sekvencii, ktorý funguje v kvardatickom čase a je založený na dynamickom programovaní. Sekvencie sú častokrát dĺžky rádovo miliardy znakov a takýto algortmus pre ne nie je vhodný. V práci sa sútreďujeme na alternatívne riešenia, ktoré bežia v lepšej časovej zložitosti a miera ich chyby je minimálna.

Kľúčové slová: dynamic time warping, minION, zarovnávanie sekvencii

Abstract

DTW (dynamic time warping) is a simple dynamic programming algorithm for sequence alignment of two sequences, which has quadratic time complexity. However, sequences are often longer than billions of characters and this algorithm is not feasible for such sequences. In this work we focus of alternative solutions which have smaller time complexity and provide a solution possibly with a small error.

Keywords: dynamic time warping, minION, sequence alignment

Contents

viii *CONTENTS*

List of Figures

x LIST OF FIGURES

List of Tables

xii LIST OF TABLES

Bibliography