



Kirjoitelma

Tietojenkäsittelytieteen kandiohjelma

Koneoppimisen menetelmät ja käyttökohteet lääketutkimuksessa -ja kehityksessä

Heikki Pulli

23.9.2021

MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
HELSINGIN YLIOPISTO

Ohjaaja(t)

Tarkastaja(t)

Yhteystiedot

PL 68 (Pietari Kalmin katu 5)
00014 Helsingin yliopisto

Sähköpostiosoite: info@cs.helsinki.fi

URL: <http://www.cs.helsinki.fi/>

Sisällys

1	Johdanto	1
2	Lääketutkimuksessa -ja kehityksessä käytetyt koneoppimisen mallit	2
3	Koneoppimismallien käyttökohteet	3
4	Koneoppimismallien heikkoudet ja puutteet	4
	Lähteet	5

1 Johdanto

Koneoppimismallien käyttö lääketutkimuksessa ja -kehityksessä on lisääntynyt merkittävästi. Lisäksi tutkimus, jossa selvitetään kuinka eri koneoppimismalleja voidaan hyödyntää lääkekehityksen tarpeisiin on lisääntynyt. Jotkin organisaatiot järjestävät kilpailuja eri tutkimusryhmien välillä, joiden tarkoituksena on yrittää löytää uusia käyttökohteita koneoppimismalleille lääketutkimuksen tutkimusalueelta. Yrityksille kannustimena käyttää koneoppimismalleja lääkkeiden kehityksessä toimii mahdollisuus vähentää lääkkeiden käytettyä aikaa ja resursseja, kun koneet voivat antaa omat ennusteensa, millä lääkkeillä on suurimmat todennäköisyydet päästä klinisiin testeihin ja niistä läpi tuotantoon. Lisäksi jotkut edistyneet mallit voivat annetusta datasta päätellä, mistä yhdisteistä muodostuu tiettyyn sairauteen tehoava lääke ja mitkä ovat tämän lääkkeen valmistusvaiheet. Nämä kone voi sitten antaa lääkkeenkehittäjille, jotka voivat validoida koneen antaman yhdisteen toimivuuden ja vaiheiden pätevyys. Nämä ovat esimerkkejä, kuinka koneoppimista voidaan hyödyntää, mutta käyttökohteita on useampia. Koneoppimisen hyödyntämisessä lääketutkimuksessa on kuitenkin vielä monia puutteita ja heikkouksia. Suurin osa puutteista liittyy datan saatavuuteen ja saatavilla olevan datan käytettävyyteen. Kaikki saatavilla oleva data ei ole aina tilannekohtaisesti tarpeeksi hyvää, jotta sitä voitaisiin käyttää haluttuun tilanteeseen. Lisäksi hyvän datan tuottamista myös hidastaa sen korkea tuotantokustannus. Lisäksi eri mallien tulosten validointia vaikeuttaa se, että ei voida koskaan täysin tietää, miten, kone on tulokseensa päätenyt. Nämä ovat ongelmia, joihin tieteellinen yhteisö yrittää saada ratkaisuja [1]. Näistä puutteista huolimatta koneoppiminen on todistanut olevansa tehokas työkalu lääkekehityksessä.

2 Lääketutkimuksessa -ja kehityksessä käytetyt koneoppimisen mallit

Lääketutkimuksessa käytetään monia eri koneoppimisen malleja. Eri mallit soveltuvat paremmin eri tilanteisiin, jolloin tutkija -tai kehittäjäryhmän tulee valita käytettävä malli ongelman mukaisesti. On kuitenkin todettu, että eri paradigmaa noudattavat mallit sopeutuvat tiettyihin tehtäviin paremmin. On huomattu, että ohjatulla oppimisella ja vahvistusoppimisella koulutetut mallit soveltuvat paremmin tehtäviin, joissa on tarkoituksena luokitella dataa tutkittavien ominaisuuksien perusteella. Näitä ovat esimerkiksi kuviin perustuva diagnoosi tai syöpään vaikuttavien geenien RNAi seulonnassa.

Ohjaamatonta koneoppimisen malleja käytetään puolestaan, kun tutkittavasta datasta halutaan löytää ryppäitä.

3 Koneoppimismallien käyttökohteet

4 Koneoppimismallien heikkoudet ja puutteet

Lähteet

- [1] J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer ja S. Zhao. "Applications of machine learning in drug discovery and development". eng. *Nature reviews. Drug discovery* 18.6 (2019), s. 463–477. ISSN: 1474-1776.

