

Treball Final de Màster

Estudi: Màster en Ciència de Dades

Títol: Modelització de les corbes d'aparició de glucosa en sang
d'àpats reals per al seu ús en aplicacions biomèdiques

Document: Resum

Alumne: Josep Noguer i Torres

Tutor: Dr. Iván Contreras Fernández-Dávila

Departament: Enginyeria elèctrica, electrònica i automàtica

Àrea: Enginyeria de sistemes i automàtica

Convocatòria: Setembre 2023

TREBALL FINAL DE MÀSTER

Modelització de les corbes d'aparició de glucosa en sang
d'àpats reals per al seu ús en aplicacions biomèdiques

Josep Noguera i Torres

TUTOR: Dr. Iván Contreras Fernández-Dávila

Màster en Ciència de Dades

UNIVERSITAT DE GIRONA

Setembre 2023

1 Introducció

L'experimentació és essencial per tal que el coneixement sigui precís, fiable i perquè aquest pugui servir de suport a l'hora de prendre decisions en una àmplia varietat de camps, provar teories, determinar l'eficàcia de tractaments o la seguretat de nous dissenys. De tota manera l'experimentació és un procés complex i requereix dissenys precisos per a reduir problemes de cost econòmic, ètica, reproductibilitat i biaix. Amb aquesta finalitat les dades sintètiques permeten representar entorns reals de manera acurada, que poden ser utilitzats per a realitzar proves i investigació. Els progressos en computació i en les diverses tècniques de generació de dades han suposat un creixement en el potencial de les dades sintètiques, ja que cada cop és possible treballar amb conjunts més grans i personalitzats que els que poden obtenir-se únicament amb dades reals. Es poden construir escenaris que d'altra manera no podrien ser simulats. Un dels camps que més es beneficia d'aquest tipus de dades és la medicina, on les dades sintètiques són una eina per a poder dur a terme proves i assaigs de noves tecnologies, protegint alhora la privacitat i la seguretat del pacient.

L'objecte d'aquest treball és el disseny i implementació d'un sistema basat en tècniques d'aprenentatge automàtic generatiu, per tal de modelar de manera acurada les corbes d'aparició de la glucosa en sang. A partir d'un set de dades d'àpats concrets es dissenya un model per a replicar las corbes que genera la ràtio d'aparició de la glucosa en sang a las hores postprandials. Les dades subministrades pel laboratori de recerca MICELab de la Universitat de Girona contenen l'especificació detallada dels diferents ingredients, informació nutricional precisa i la ràtio d'aparició de glucosa exògena (RA) en sang. Es busca obtenir un sistema basat en dades, capaç de millorar l'aproximació dels models metabòlics actuals, que funcionen amb mètodes matemàtics i la principal debilitat dels quals és la seva naturalesa determinista que els inhabilita per a determinades tasques que requereixen de la variabilitat pròpia d'un entorn real. .

2 Metodologia

La metodologia proposada es basa en arquitectures d'aprenentatge profund (C-GAN, *autoencoders*, *transformers*, entre d'altres) per a millorar les aproximacions actuals a partir de l'aprenentatge de les distribucions de dades reals, condicionades als paràmetres que defineixen cada àpat (ingredients, valor nutricional, etc.). Concretament es fan servir xarxes del tipus generativa adversària condicionada que després de l'estudi de la viabilitat del projecte s'han considerat les més adients degut a la petita quantitat de mostres disponibles. El model implementat és una adaptació de les xarxes tipus *Pix2Pix* però amb les modificacions pertinents segons la naturalesa temporal del problema que es planteja en aquest projecte. Per a atacar els inconvenients característics d'aquest tipus de xarxa s'ha adaptat el model seguint les indicacions del model GAN de *Wasserstein* fins a

assolir un entrenament correcte.

Per a poder alimentar el model ha sigut necessari aplicar un seguit de transformacions a les dades originals i determinar quines eren les variables més rellevants per a que el model pogués condicionar correctament les corbes. Aquestes han estat per una banda la quantitat d'hidrats de carboni que s'han normalitzat a valors entre 0 i 1 i per altra el tipus d'aliment. Aquesta segona variable, que originalment distingia els àpats en quatre categories (*Small*, *Fast*, *Medium* i *Large*) s'ha agrupat solament en dues (*Fast* i *Non-Fast*) ja que el fet diferencial entre els dos subgrups de cada categoria és els hidrats de carboni, dada de la que ja disposem al model. Les categories s'han adaptat per a treballar en una codificació tipus *one-hot*

3 Resultats

Els resultats son contrastats a través de la validació estadística pertinent i comparant corbes reals amb les generades a partir de noves variables condicionants seguint la mateixa distribució de valors que al set de dades original. S'han dissenyat un seguit de mètriques i experiments que avaluen tant la validesa de les corbes generades com la relació entre aquestes i les diferents variables que la condicionen.

Les mètriques que s'han definit són el temps en que la corba arriba al primer pic, l'àrea sota la corba de la RA i el pendent de la corba després del pic. Amb això els resultats obtinguts són els que es mostren a la taula 1. En aquests es presenta la comparació entre les mostres sintètiques i reals per a cada classe d'àpat (Fast i Non-Fast)

	Temps fins al pic	Àrea sota la corba	Pendent després de pic
Fast sintètic	5.96(3.11)	52.22(23.51)	4.88(5.02)
Fast real	7.39(5.81)	48.89(41.48)	4.09(5.34)
Non-Fast sintètic	12.94(4.46)	2246.36(402.86)	3.01(2.9)
Non-Fast real	15.03(9.54)	2509.47(1526.31)	2.4(3.25)

Table 1: Comparació dels valors de *mitja* (*desviació estàndard*) entre els dos tipus de mostres generades i reals.

	Temps fins al pic	Àrea sota la corba	Pendent després de pic
Fast sintètic	6.0(4.0 – 7.0)	48.37(35.57 – 66.28)	3.4(0.48 – 7.7)
Fast real	5.5(3.0 – 8.75)	32.02(18.16 – 75.12)	2.84(–0.0 – 4.4)
Non-Fast sintètic	13.0(9.0 – 15.0)	2199.63(1951.41 – 2495.81)	2.2(0.7 – 4.46)
Non-Fast real	14.0(8.75 – 18.25)	2194.31(1457.66 – 3134.72)	0.96(–0.08 – 4.18)

Table 2: Comparació de la distribució *mediana* ($q1$ - $q3$) entre els dos tipus de mostres generades i reals.

4 Conclusions

Tal i com es mostra en l'apartat de resultats els valors que s'han obtingut demostren per una banda que el model és capaç de mantenir els estadístics propis del les sèries de la RA originals. També a partir dels altres experiments realitzats es pot veure com les variables que condicionen el model fan aparèixer les variacions esperades en el model, podent veure quina és la influència tant de la classe d'àpat com de la quantitat d'hidrats de carboni.

A més, aquest estudi ha destacat la importància de la utilització de models generatius profunds en la recerca biomèdica, ja que permeten explorar i comprendre millor les dinàmiques de la glucosa exògena en sang, la qual cosa pot tenir implicacions significatives en el tractament i control de la diabetis i altres malalties relacionades.

En última instància, els models generatius profunds es perfilen com una eina prometedora en la generació de corbes de glucosa exògena, la qual cosa podria contribuir a millorar l'atenció mèdica i el disseny de teràpies personalitzades en el futur. Tanmateix, és important destacar que aquest camp de recerca encara presenta desafiaments i àrees de millora, com la necessitat de dades d'alta qualitat i l'optimització de la precisió dels models. En resum, aquest treball estableix les bases per a futures investigacions en l'ús de models generatius profunds en el camp de la glucosa exògena en sang, amb l'objectiu de millorar la qualitat de vida de les persones afectades per trastorns metabòlics.