

DiabManager

Projet d'ingénierie informatique - Rapport final



Groupe 3

Mathis DELEHOUZÉE

Gustavo MAGAÑA LÓPEZ



Dirigé par Monsieur
Mohammed BENJELLOUN

Année académique 2018-2019

Remerciements

Merci à

Mons, le 28 avril 2019

Table des matières

Remerciements	1
Abréviations	3
Introduction	4
1 Contexte et motivation du projet	5
1.1 Le diabète	5
1.2 Le besoin d'un système de monitoring et de prédiction	5
1.3 Difficultés dans l'élaboration d'un algorithme prédictif du taux de glucose	5
2 État de l'art	5
2.1 Méthode d'analyse temporelle	5
2.1.1 Auto-regression	5
2.1.2 Modèle ARIMA	5
2.1.3 Prédiction par des médecins	5
2.2 Méthode de machine learning	7
2.2.1 Réseau de Neurone	7
2.2.2 Support vector regression	7
2.3 Discussion sur l'ajout d'un modèle physiologique	7
3 Cahier des Charges	7
3.1 Outils utilisés	7
3.2 Objectif 1 : Modèles de classification	7
3.2.1 Support vector machine	7
3.2.2 Discussions sur les autres méthodes de classification	7
3.3 Objectif 2 : Modèle de régression	7
3.3.1 Support vector regression	7
4 Solutions et résultats	7
4.1 Résultats du SVM	7
4.2 Passage du SVM au SVR	7
4.3 Résultat du SVR	8
5 Perspectives d'améliorations	8
5.1 Ajout de certains paramètres physiologique non contraignant au SVR	8
Conclusion	8
Annexes	8

Abréviations

SVR : Support vector regression

SVM : Support vector machine CGM : Continuous glucose monitoring CGM : Continuous glucose monitoring BG : Blood glucose T1D : Diabète de type 1

Introduction

Dans le cadre de notre cours d'ingénierie informatique, il nous a été demandé de réaliser un projet dans le domaine de l'IT. Malgré les nombreux projets intéressants proposés, nous avons décidé de réaliser un projet personnel portant sur le diabète de type 1.

Le diabète de type 1 (T1D) est une maladie chronique touchant de plus en plus de personnes à travers le monde.

Il s'agit d'une déficience en insuline (le patient est nommé insulino-dépendant). Cette molécule permet d'assimiler le sucre dans notre organisme afin d'apporter à nos organes l'énergie nécessaire à leur bon fonctionnement.

Le patient doit donc régulièrement contrôler son taux de glucose sanguin, blood glucose (BG) en anglais, afin de ne pas avoir un taux anormal néfaste pour son organisme.

En effet, un taux de glucose sanguin supérieur à 150 mg/dL indique qu'une personne est en hyperglycémie. À l'inverse un BG inférieur à 70 mg/dL signifie que la personne est en hypoglycémie. Ces différents états doivent être à tout prix évité par le patient car ils causent généralement des séquelles à long terme. L'objectif est donc d'optimiser ses prises de glucide et d'insuline grâce à un système de monitoring.

Quand une personne est diagnostiquée diabétique, un nouveau style de vie démarre pour lui. À partir du diagnostic le patient devra commencer à estimer la quantité de glucides dans les aliments qu'il ingère et s'injecter une dose d'insuline en conséquence qui lui permettra de bien métaboliser le sucre. Cet exercice demande beaucoup de rigueur et de réflexion. Notre projet vise donc à améliorer la qualité de vie des diabétiques en facilitant leur gestion de leur taux glycémique grâce à la mise en place d'un algorithme prédictif annonçant un état dangereux pour le patient.

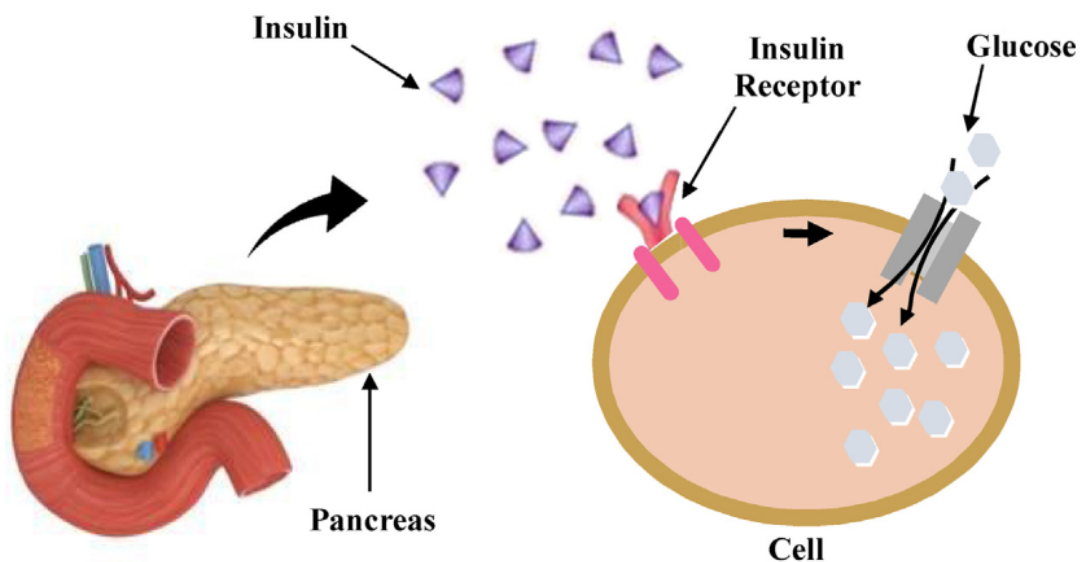


FIGURE 1 – Processus de l'absorption du glucose grâce à l'insuline

1 Contexte et motivation du projet

1.1 Le diabète

1.2 Le besoin d'un système de monitoring et de prédiction

1.3 Difficultés dans l'élaboration d'un algorithme prédictif du taux de glucose

Le problème est qu'il n'est pas toujours évident de connaître la quantité d'insuline à s'injecter précisément du fait que le taux de sucre dépend de nombreux facteurs (alimentation, activité physique, taux des dernières heures, etc.). Pour rester en bonne santé, les médecins préconisent de vérifier son taux régulièrement et de s'injecter de l'insuline afin de garder un taux stable. À cette fin, nous avons décidé de réaliser un algorithme prédictif permettant de faciliter la vie de nombreuses personnes. Cet algorithme permettra de donner une prédiction de l'évolution du taux d'insuline en fonction du taux actuel et d'une variable appliquée, c'est à dire, un événement indiqué par l'utilisateur (sport, alimentation,...). Elle donnera également à l'utilisateur la quantité d'insuline à s'injecter et des conseils pratiques de médecins afin d'améliorer son contrôle glycémique.

Pour résoudre cette problématique nous sommes aidé par la Dr. Isabelle Paris, endocrinologue Diabétologue au CHR Mons-Hainaut.

La Dr. Paris nous a mis au courant des nombreuses applications existantes et va nous dresser une liste non-exhaustives des dernières innovations dans le domaine afin que nous puissions nous en inspirer.

2 État de l'art

2.1 Méthode d'analyse temporelle

2.1.1 Auto-regression

2.1.2 Modèle ARIMA

Il existe de nombreux moyens pour essayer de prédire un état glycémique. Les recherches actuelles se concentrent sur des prédictions de l'ordre de 30' à 1h afin de permettre aux patients d'avoir un laps de temps convenable pour réagir. Ce que nous avons retenus des différents articles utilisés pour nous guider dans ce projet est que la plupart des algorithmes issus du machine learning sont plus performants que les prédictions des médecins spécialistes. Ce constat nous rassure sur notre démarche et nous pousse à nous renseigner d'autant plus sur le sujet.

Dans notre état de l'art, nous exposerons différents travaux et résultats sur lesquels nous avons batis notre raisonnement et nos expériences.

2.1.3 Prédiction par des médecins

L'article de l'université de l'Ohio sur l'approche du machine learning pour la prédiction du glucose sanguin [1] nous donne une bonne mesure sur la précision des prédictions données par des médecins spécialistes.

Horizon	t_0	ARIMA	Phys ₁	Phys ₂	Phys ₃
30 min	27.5	22.9	19.8	21.2	34.1
60 min	43.8	42.2	38.4	40.0	47.0

FIGURE 2 – Erreur quadratique moyenne des prédictions des différentes méthodes

La colonne horizon fait référence à l’horizon de prédiction soit 30 ou 60 minutes. Le tableau nous permet donc de comparer les erreur quadratique moyenne (RMSE) de certaines méthode. La première méthode, t_0 sert de point de départ pour des comparaisons et soutien le fait que le taux de glucose ne change jamais. On peut donc forcément s’attendre à ce que cette méthode hasardeuse puisse avoir une RMSE élevée.

La méthode suivante est appelée Auto-Regressive Integrated Moving Average (ARIMA) (Box, Jenkins, and Reinsel 2008) et correspond à une modélisation de série chronologique (méthode autorégressive à moyenne mobile intégrée). Il s’agit d’une méthode déterminant un modèle sous-jacent pouvant prédire l’évolution du modèle en donnant un sens général au mouvement des données. Un gros défaut de cette méthode est qu’elle n’est utilisable que sur des données stationnaires et donc que le procédé ne possède aucune tendance haussière ou baissière. La moyenne du procédé doit donc être particulièrement stable ce qui n’est pas forcément le cas des jeux de données de l’étude. Ce qui explique donc la faible qualité de la méthode afin de prédire les BG futurs.

Ces résultats sont à comparer avec ceux de 3 docteurs, spécialistes du diabète,

Le test de prédiction des médecins s’est effectué sur 5 patients T1D avec un ensemble de données d’évaluation de 200 points (40 par patient). Les médecins avaient accès aux données précédentes du patient afin de pouvoir composer leur propre généralisation de la fluctuation du glucose sanguin. Après cela, il leur été demandé d’effectuer des prédictions pour 30 et 60 minutes. Après chaque prédiction la véritable valeur du BG leur été fournie afin qu’ils puissent affiner leur réflexion.

Plus loin dans l’article, un nouveau test est effectué mais cette fois ci grâce à une méthode de machine learning : le support vector regression (SVR). Comme nous le voyons sur la figure 3, le résultat est sans appel. Cette dernière méthode est bien plus performante que les 2 méthodes précédentes et que le meilleur résultat des médecins. Nous détaillerons le principe de la méthode SVR plus loin dans notre travail.

Horizon	t_0	ARIMA	Phys ₁	SVR ϕ	SVR ϕ +A
30 min	27.5	22.9	19.8	19.6	19.5
60 min	43.8	42.2	38.4	36.1	35.7

FIGURE 3 – Erreur quadratique moyenne des prédictions des différentes méthodes

Il est à noter que le dernier cas présent sur la figure 3 est le cas d’un modèle SVR prenant en considération des facteurs physiologiques tel que le stress, la fatigue, la sensibilité à l’insuline, ... Même si ce dernier semble plus performant avec une RMSE légèrement plus faible, nous souhaitons opter pour un modèle prédictif non intrusif ou contraignant pour le patient. En effet, un tel modèle demande des entrées régulières de certaines données peu objectives ou quantifiables par le patient (Humeur, maladie,...) et ne constitue donc pas une voie exploitable pour l’amélioration du quotidien des personnes diabétiques.

Cette première approche des possibilités de réalisation d’algorithmes prédictif nous amène à considérer la méthode SVR très sérieusement pour nos propres recherches. En effet, cette méthode basée sur la lecture de données antérieures afin de créer sa prédiction est très peu contraignante pour la patient mais reste tout de même à un bon niveau de performance.

2.2 Méthode de machine learning

2.2.1 Réseau de Neuronne

2.2.2 Support vector regression

2.3 Discussion sur l'ajout d'un modèle physiologique

Unfortunately, these methods require the patient to specify their daily activities : meal intake, insulin injection and emotional factors, which can be error prone.

3 Cahier des Charges

- Principal : Modèle de classification

Même avec la formation donnée par les médecins, diététiciens, etc il existe toujours une certaine incertitude chez le patient : Est-ce que cette combinaison de facteurs va amener mon taux de sucre vers le bas ou vers le haut ? Dit autrement : Ce que je veux manger ou cette activité physique que je suis en train de pratiquer fera-t-elle en sorte que je tombe en hypo ou hyperglycémie ? Les modèles de prédiction intégrés dans les pompes à insuline ne prennent en compte que le taux actuel du sucre et la tendance.

Afin d'améliorer la valeur de notre algorithme, surtout au niveau thérapeutique, la prédiction se fera sous la forme d'un modèle de classification plutôt que de régression. En fonction des facteurs actuels, pour chaque patient.

- Secondaire : Modèle de prédiction

Si nous arrivons à entraîner un modèle assez efficace, nous aimerions passer à un modèle qui donnera au patient une valeur numérique de la concentration du sucre dans le sang en prenant en compte les facteurs externes précédemment mentionnés.

3.1 Outils utilisés

Dans un premier temps nous avons décidé de créer un GitHub afin de faciliter le travail collaboratif :

<https://github.com/gmagannaDevelop/DiabManager.git>

Pour compiler notre code Python, nous utilisons Jupyter programme très efficace pour notre projet. Cela nous permet de voir les graphiques et résultats pour chaque étape du code.

3.2 Objectif 1 : Modèles de classification

3.2.1 Support vector machine

3.2.2 Discussions sur les autres méthodes de classification

3.3 Objectif 2 : Modèle de régression

3.3.1 Support vector regression

4 Solutions et résultats

4.1 Résultats du SVM

4.2 Passage du SVM au SVR

Ici on explique la différence entre le SVR et le SVC ...

Pour construire un bon modèle de SVR il nous faut d'abord effectuer une recherche en grille (grid search) afin de trouver les paramètres γ et C qui minimiseraient les erreurs.

À titre de comparaison voici les résultats obtenus dans d'autres recherches [2]

- Regression type epsilon SVR
- Radial basis function (RBF) as kernel
- ϵ set to default value of 0.001
- (Kernel coefficient) γ set to a range from $[2^{-15}, 2^{-14}, \dots, 2^3]$
- (Penalty parameter) C set to a range from $[2^{-5}, 2^{-4}, \dots, 2^{15}]$

FIGURE 4 – Paramètres déterminés par grid search dans l'étude de l'IEEE 2016

4.3 Résultat du SVR

Afin de déterminer la valeur optimale pour notre modèle SVR, nous avons décidé d'effectuer une grid search. Voici les résultats obtenus pour notre modèle avec et sans grid search.

5 Perspectives d'améliorations

5.1 Ajout de certains paramètres physiologique non contraignant au SVR

Mettre ici la GUI qu'on a faite

Conclusion

"Diabetes management presents many opportunities for clinicians and artificial intelligence researchers to collaborate on challenging research that could potentially improve health, safety and quality of life for millions of people living with diabetes"¹

Annexes

[3]

Références

- [1] C. M. J. S. Kevin Plis, Razvan Bunescu and F. Schwartz, "A machine learning approach to predicting blood glucose levels for diabetes management," *Ohio University*, p. 5, 2014.
- [2] B. H. G. C. M. P. B. Maximilian P. Reymann, Eva Dorschky and B. M. Eskofier, "Blood glucose level prediction based on support vector regression using mobile platforms," *IEEE, EMBS*, pp. 2990–2993, 2016.
- [3] S. M. Pappada, B. D. Cameron, and P. M. Rosman, "Development of a neural network for prediction of glucose concentration in type 1 diabetes patients," *Journal of Diabetes Science and Technology*, vol. 2, no. 5, pp. 792–801, 2008.

1.