Optimal Insulin Dosing for Glucose Control in a Virtual Type-I Diabetes Patient through Reinforcement Learning

Abhishek Agarwal, Aditya Pareek, Vishnu Masampally, Venkataramana Runkana*

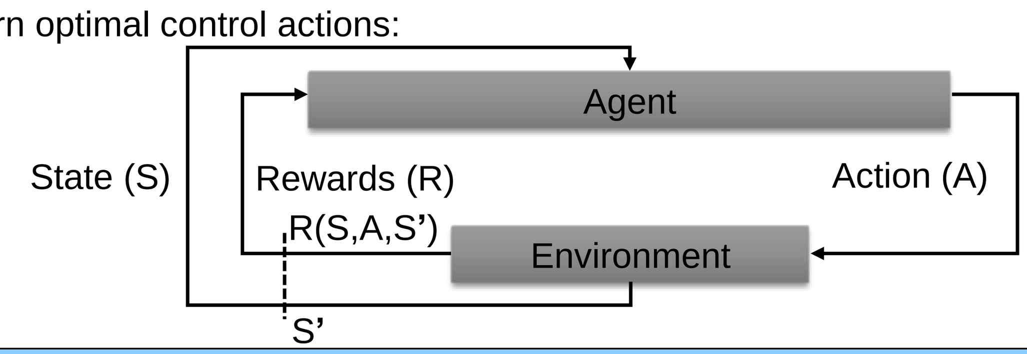
TCS Research, Tata Research Development and Design Centre,
Tata Consultancy Services, 54B, Hadapsar Industrial Estate, Pune - 411013, India
***e-mail: venkat.runkana@tcs.com**

Introduction

- Type 1 Diabetes Mellitus (T1DM): The body's immune system destroys β-cells, eliminating insulin production from the body
- T1DM patient depends on the exogenous insulin dosages
- Open loop control comprising multiple daily insulin injections generally leads to poor glycaemic control
- Closed loop control using an Artificial pancreas device system (APDS) is desired. APDS consists of:
  - Continuous glucose monitoring sensor (CGMS)
  - Controller that estimates insulin to be dosed based on glucose and other measurements
  - Insulin pump

Reinforcement learning and its application in APDS

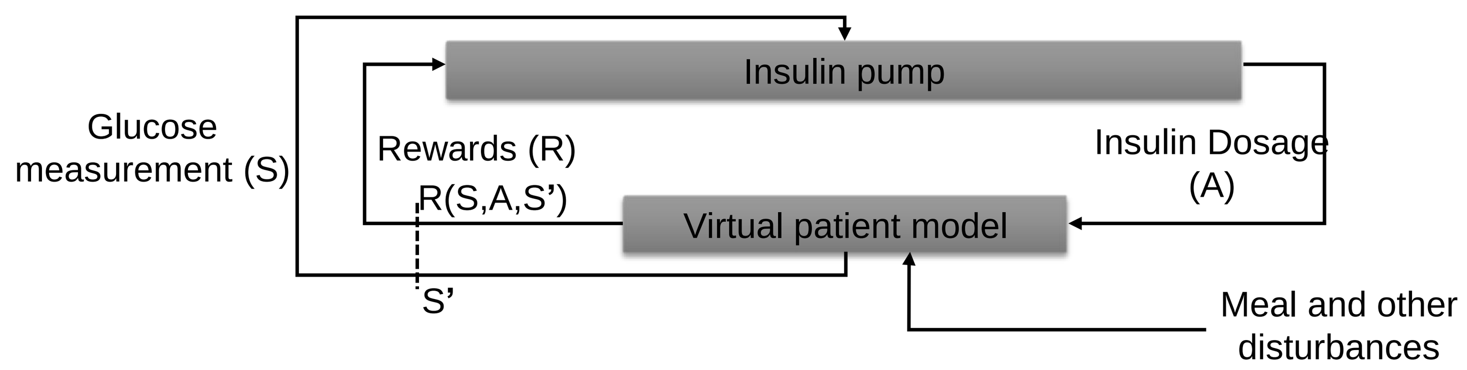
- Uncertainty associated with external disturbances such as amount of meal, physical activity[1,2]
- Inter and intra patient variability in glucose metabolism
- RL framework can account for such unexpected disturbances and individualize insulin dosing
- RL algorithms that can help agent learn optimal control actions:
  - Model-based
  - Model-free algorithms
    1. Q-learning
    2. Dyna-Q
    3. Q(λ)



Mathematical model of a virtual diabetes patient[3]

- This model is used previously for in-silico testing of various control algorithms
- Glucose enters via:
  - Intestinal absorption through meals
  - Hepatic glucose production
- Glucose is removed via:
  - Utilization in RBC
  - Insulin-dependent glucose utilization in the liver
  - Glucose excretion takes place above the renal threshold
- In T1DM patient, only source of insulin is through APDS.
- Coupled model has 3 ordinary differential equations and 4 algebraic equations

Closed loop Type-1 glucose control problem



- Reward function:

$$R_{t+1} = \begin{cases} -\alpha_1 + -\alpha_2(S_{t+1} - S_t), & G_t < 3.8 \\ \beta, & 3.8 < G_t < 7.6 \\ -\alpha_1 + -\alpha_2(S_{t+1} - S_t), & G_t > 7.6 \end{cases}$$

Element	Detail
State space	Plasma glucose (G): {2, 2.2, 2.4, ……, 20} Meal: {ON, OFF}
Action space	u_{ins} = {0, 0.05, ……,,0.5} mU/min

Online RL control algorithms

- **Q-learning:**

$$Q(S_t, A_t) = Q(S_t, A_t) + \alpha\left[R_{t+1} + \gamma \underset{a}{arg\,max}\, Q(S_{t+1}, a) - Q(S_t, A_t)\right]$$

$$\pi_t(S_t) = \begin{cases} P\left(\underset{a}{arg\,max}\, Q(S_t, a)\right) = 1 - \epsilon \\ P\left(A \neq \underset{a}{arg\,max}\, Q(S_t, a)\right) = \epsilon/(N_A - 1) \end{cases}$$

- **Q(λ):**

Initialize $Q(s,a)$ arbitrarily, for all $s \in \mathcal{S}$, $a \in \mathcal{A}$
Repeat (for each episode)
  $Z(s,a) = 0$ for all $s \in \mathcal{S}$, $a \in \mathcal{A}$
  Initialize S, A
  Repeat (for each step of episode)
    Take action A, observe R, S'
    Chose A' at S' using policy derived from Q
    $A^* = argmax_a Q(S', a)$
    $\delta = R + \gamma Q(S', A') - Q(S, A)$
    $Z(S, A) = Z(S', A') + 1$
    For all $s \in \mathcal{S}$, $a \in \mathcal{A}$
      $Q(s,a) = Q(s,a) + \alpha\delta Z(s,a)$
      If $A' = A^*$, then
        $Z(s,a) = \gamma\lambda Z(s,a)$
      else $Z(s,a) = 0$
        $S = S'; A = A'$
  Until S is terminal`

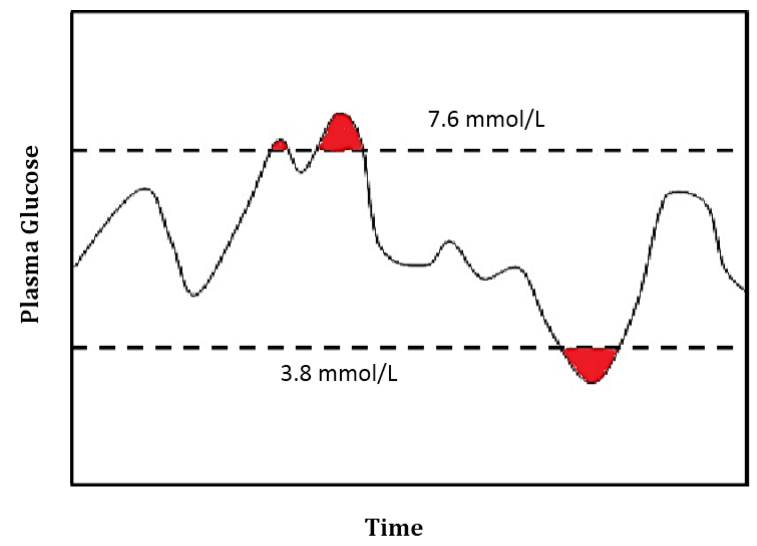
- **Dyna-Q:**

Initialize $Q(s,a)$ and $Model(s,a)$ arbitrarily, for all $s \in \mathcal{S}$, $a \in \mathcal{A}$
Repeat (for each episode)
  Initialize S, A
  Repeat (for each step of episode)
  Chose A at S using policy derived from Q
  Observe R, S'
  Update Q
  $Q(s,a) = Q(s,a) + \alpha\left(R + \gamma argmax_a Q(S', a) - Q(S, A)\right)$
  Update $Model$
  $Model(S, A) = R, S'$
  Repeat N number of times
    S = select a state at random from the visited states
    A = select a state at random from the visited states
    Sample $Model(S, A)$
    $R, S' = Model(S, A)$
    Update Q
    $Q(s,a) = Q(s,a) +$
      $\alpha\left(R + \gamma argmax_a Q(S', a) - Q(S, A)\right)$
 Until S is terminal

- Gaussian disturbances in:
  1. Amount of meal: 30% of the mean
  2. Time of meal: 60 min
- Control signal to the pump is updated every 10 minutes

Performance measure:

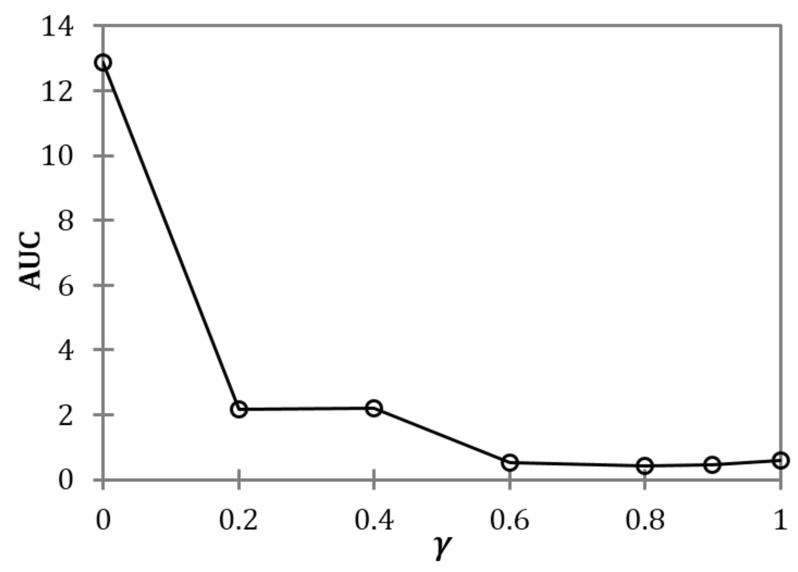
- Area under Curve (AUC) enclosed by continuous glucose measurements with its upper and lower bounds of permissible limits during a 24 hour period is calculated.
- Lower AUC signifies better control.
- Average of AUC over all episodes (5000) is used to compare the performance of different algorithms

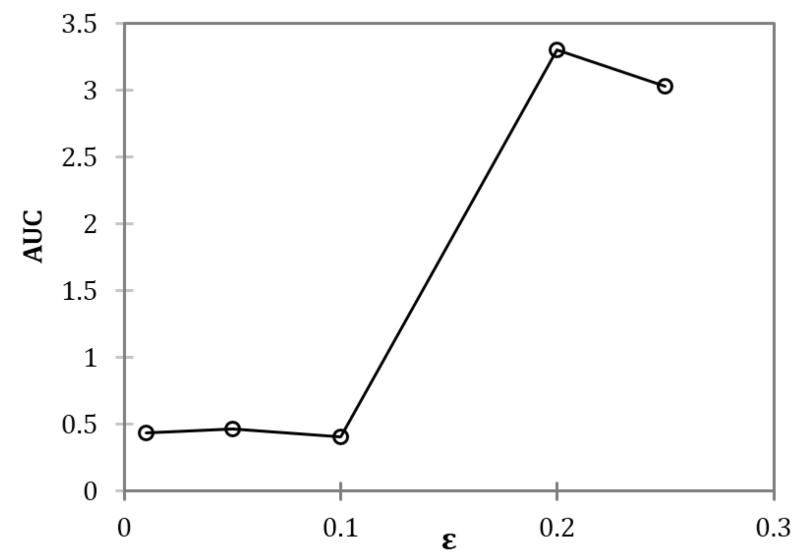


Results

Effect of tuning parameters of RL algorithm, namely, α, γ, and ε on the quality of control:

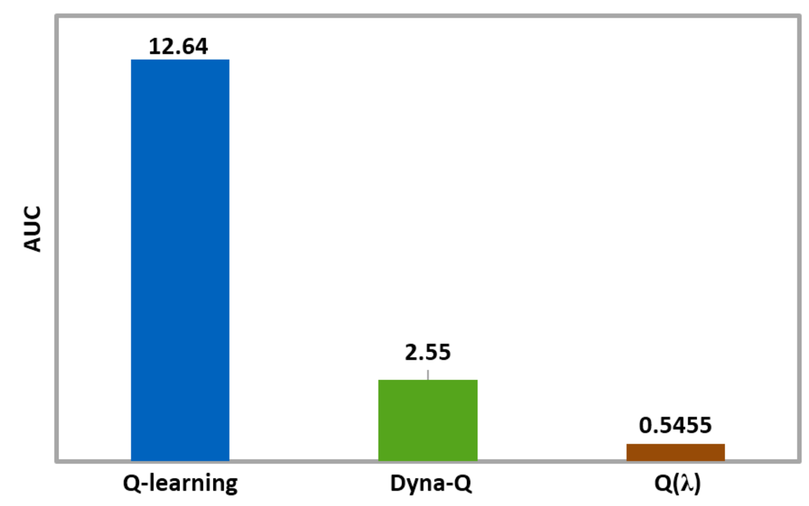
- Learning rate, α, above 0.2 yielded similar glucose control in the 5000 episodes simulated
- AUC reduced significantly on increasing discount factor γ and became steady at 0.6
- An ε of 0.1 was found to minimize the area outside the desired glucose range



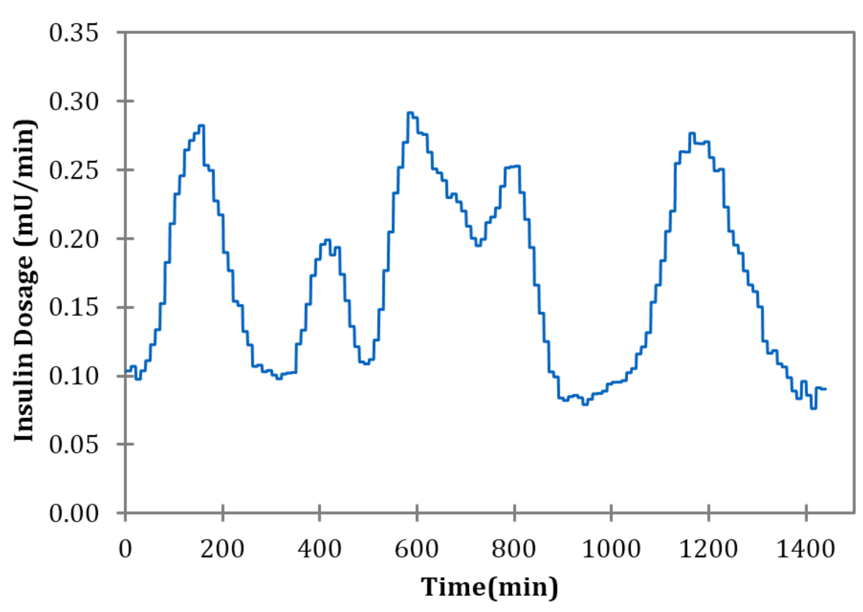


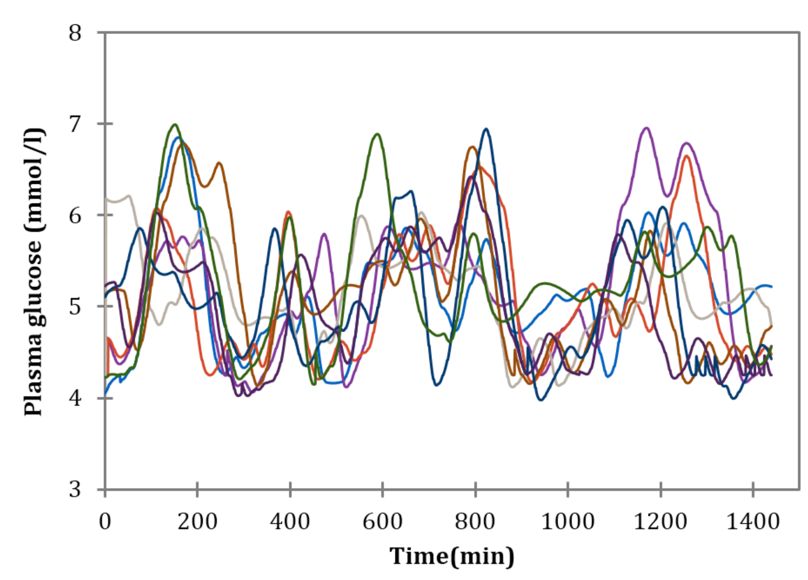
Performance comparison of RL algorithms:

- Dyna-Q which incorporates a planning agent performs better than Q-learning
- Q(λ) outperforms both the algorithms
- Dyna-Q performance can be further improved by increasing the number of planning steps but this comes with an additional computational cost
- Q(λ) is preferred over the other two as it performs better at relatively lower computational cost.



Application of Q(λ) algorithm:





- RL agent is able to control the glucose concentration during most of the days.
- Average glucose concentration is maintained between 4.5 to 5.5 mmol/L
- Peaks in the insulin dosage plot coincides with the meal intake

Conclusion

- Reinforcement learning is a viable alternative to traditional controllers used in APDS.
- In the present study, application of RL based controllers was found to be effective in maintaining normoglycemia even in presence of disturbances in meal related inputs.
- More studies need to be performed to study the efficacy of RL based controllers when intra-day variability in a virtual-patient is present.
- Moreover, RL based controllers need to be supported with either estimation methods or clinical heuristics/rules to avoid any exploratory action that can be damaging to the patient

References

1. Bequette, B. (2012). Challenges and recent progress in the development of a closed-loop artificial pancreas. *Annual Reviews in Control*, 36(2), pp.255-266.
2. Bothe, M., Dickens, L., Reichel, K., Tellmann, A., Ellger, B., Westphal, M. and Faisal, A. (2013). The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas. *Expert Review of Medical Devices*, 10(5), pp.661-673.
3. Lehmann, E.D. and Deutsch, T., 1992. A physiological model of glucose-insulin interaction in type 1 diabetes mellitus. *Journal of biomedical engineering*, 14(3), pp.235-242.
4. Acikgoz, S.U. and Diwekar, U.M., 2010. Blood glucose regulation with stochastic optimal control for insulin-dependent diabetic patients. *Chemical Engineering Science*, 65(3), pp.1227-1236

TATA
TATA CONSULTANCY SERVICES

Acknowledgement
Authors would like to thank Mr. K Ananth Krishnan, CTO, Tata Consultancy Services for his constant encouragement and support during this project.