به نام خدا

تمرین پنجم درس شبکه های عصبی توحید عابدینی ۹۹۷۲۲۱۵۳

سوال ۱)

بخش a)

برای حل مساله حالات زیر را تعریف میکنیم:

بر اساس مقدار میگو های باقی مانده به صورت کمی چهار حالت empty – low – medium – high را خواهیم داشت. که همانطور که از نام حالات مشخص هست بیانگر میزان میگو های باقی مانده هستند. (خالی – کم – متوسط – زیاد)

اعمالی که میتوانیم انجام دهیم شامل اعمال زیر است:

sale: صید میگو در این ماه انجام شود. (به صورت نسبی و بخشی از میگو های موجود صید شوند.)

not_to_sale: صید میگو در این ماه انجام نشود.

در حالت خالی هم یک عمل به اسم re-breed داریم که هزینه بالایی دارد و باعث تجدید جمعیت میگو ها میشود.

احتمالات گذار به صورت زیر هستند:

فرمت کلی ما اینگونه است :

$T(S, a, S') = R\ ,\qquad S = current\ state, a = action, S' = next\ state\ , R = reward$

$T(Empty, Re-breed, Low) = 1.0$

$T(Low, Sail, Low) = 0.25$
$T(Low, Sail, Empty) = 0.75$
$T(Low, Sail, Low) = 0.25$
$T(Low, Not-to-sale, Low) = 0.3$
$T(Low, Not-to-sale, Medium) = 0.7$

$T(Medium, Sail, Low) = 0.75$
$T(Medium, Sail, Medium) = 0.25$
$T(Medium, Not-to-sale, Medium) = 0.25$
$T(Medium, Not-to-sale, High) = 0.75$

$T(High, Sail, Medium) = 0.6$
$T(High, Sail, High) = 0.4$
$T(High, Not-to-sale, High) = 0.95$
$T(High, Not-to-sale, Medium) = 0.05$

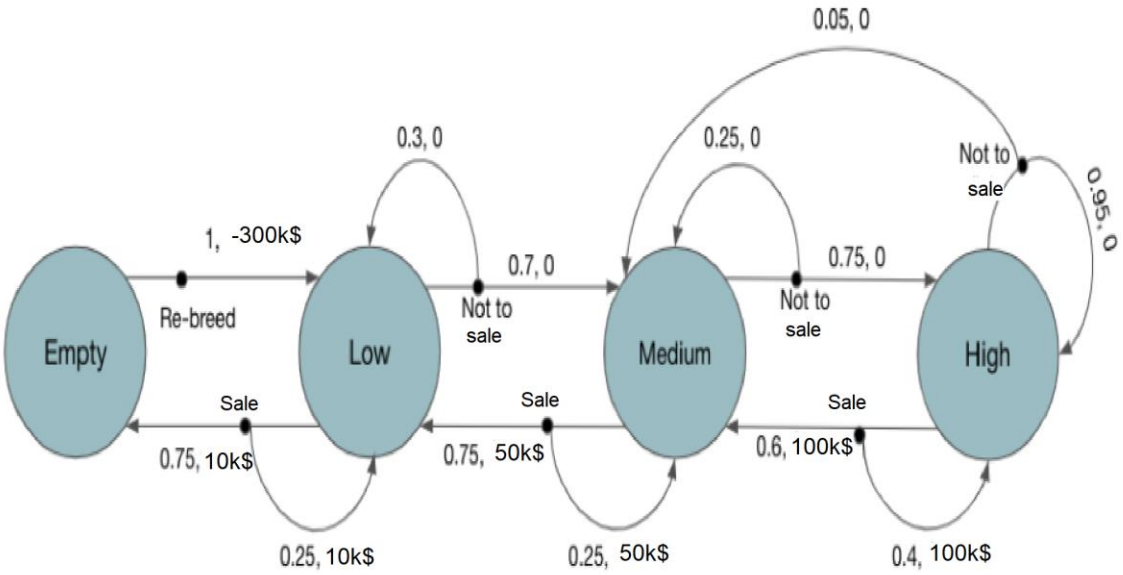
پاداش های این مساله به این صورت تعریف میشوند:

پاداش ماهیگیری در حالت های low – medium – high به ترتیب ۱۰ و ۵۰ و ۱۰۰ (واحد پول – فرض میکنیم هزار دلار) هستند.

اگر عملی ما را به حالت بدون میگو ببرد (empty) منفی ۳۰۰ هزار دلار پاداش خواهد داشت.

در نهایت مدل MDP ما به صورت زیر خواهد بود:

بدیهی است در حالاتی که صید داریم احتمال اینکه به یک حالت با مقدار میگو کمتر برویم بیشتر از این است که در حالتی با همان مقدار نسبی میگو بمانیم. (و همینطور برای عمل عدم صید هم این اصل منطقی به صورت بر عکس برقرار است.)



منابع:

https://towardsdatascience.com/real-world-applications-of-markov-decision-process-mdp-a39685546026

بخش b)

فرض میکنیم در ابتدا سرعت حرکت ماریو صفر است.

برای حل مساله حالات زیر را با مفروضات گفته شده تعریف میکنیم:

بر اساس اینکه در چه مکانی از بازی هستیم و ماریو زنده است یا نه حالات را تعریف میکنیم.

فرض میکنیم در هر منطقه به صورت احتمالی با حرکت یا پرش کند با احتمال بیشتری در آن باقی خواهیم ماند و اگر حرکت یا پرش تندی داشته باشیم با احتمال بیشتری به منطقه دیگر خواهیم رفت. (برای راحت تر کردن مساله – بدیهی است مساله به سادگی مدل داده شده نیست و برای ساده تر کردن آن فرض هایی در نظر گرفته شده است.)

پس حالات ما عبارت اند از:

left_side: ماریو در سمت چپ بازی است.

platform: ماریو در منطقه وسط بالای چاله است.

pit: ماریو در چاله است.

pipe: ماریو بر روی لوله است.

right_side: ماریو در سمت راست بازی است.

goal: ماریو به هدف رسیده است.

killed: ماریو کشته شده است. (در بقیه حالات در تمامی موارد ماریو زنده است. این حالت از pit و pipe ممکن است منتج شود و برای همین یک حالت جدا در نظر گرفته شده است)

اعمالی که میتوانیم انجام دهیم شامل اعمال زیر است:

اعمال ما به چهار عمل حرکت کند – حرکت تند – پرش کند – پرش تند تقسیم میشوند. (یعنی کندی یا تندی به صورت قابل تنظیم با استفاده از اعمال در نظر گرفته شده اند.)

fast_move - slow_move - fast_jump - slow_jump

احتمالات گذار به صورت زیر هستند:

فرمت کلی ما اینگونه است :

$T(S, a, S') = R\ ,\qquad S = current\ state, a = action, S' = next\ state\ , R = reward$

$T(Left\_side, fast\_jump, Platform) = 0.8$
$T(Left\_side, fast\_jump, Pit) = 0.2$
$T(Left\_side, fast\_move, Pit) = 0.8$
$T(Left\_side, fast\_move, Left\_side) = 0.2$
$T(Left\_side, slow\_jump, Left\_side) = 0.9$
$T(Left\_side, slow\_jump, Pit) = 0.1$
$T(Left\_side, slow\_move, Left\_side) = 0.9$
$T(Left\_side, slow\_move, Pit) = 0.1$

$T(Platform, fast\_jump, Right\_side) = 0.7$
$T(Platform, fast\_jump, Pipe) = 0.3$
$T(Platform, fast\_move, Pit) = 0.9$
$T(Platform, fast\_move, Platform) = 0.1$
$T(Platform, slow\_jump, Platform) = 0.3$
$T(Platform, slow\_jump, Pipe) = 0.5$
$T(Platform, slow\_jump, Pit) = 0.2$
$T(Platform, slow\_move, Left\_side) = 0.9$
$T(Platform, slow\_move, Pit) = 0.1$

$T(Pit, any\_action, Killed) = 1$

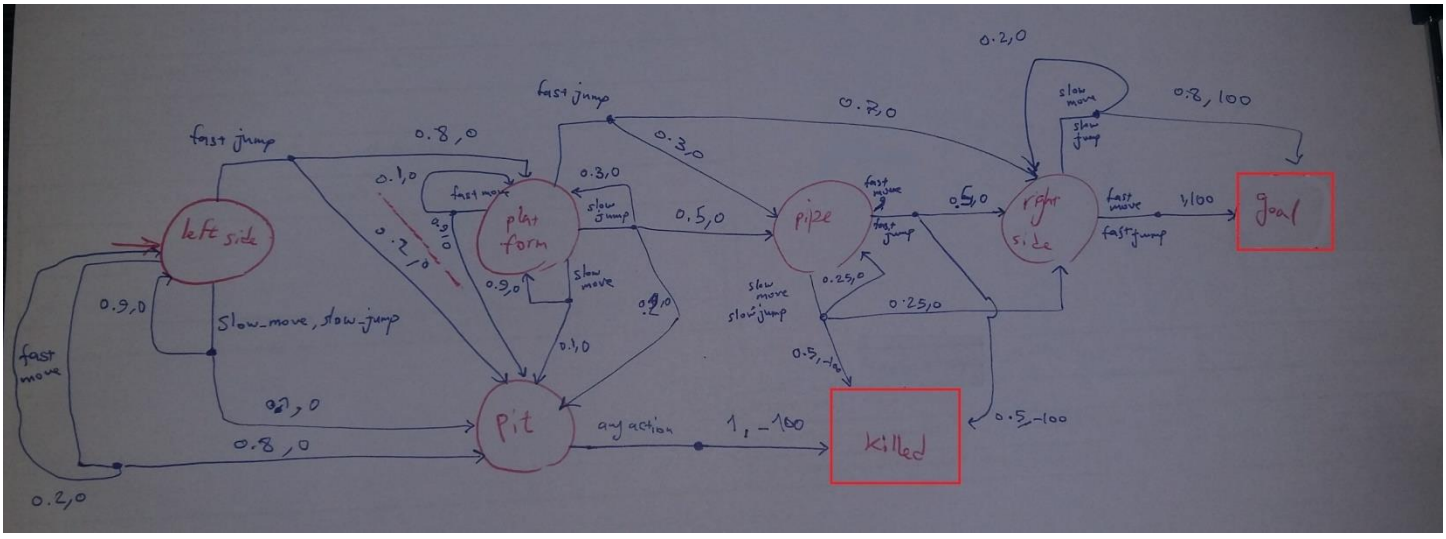
$T(Pipe, fast\_jump, Right\_side) = 0.5$
$T(Pipe, fast\_jump, Killed) = 0.5$
$T(Pipe, fast\_move, Right\_side) = 0.5$
$T(Pipe, fast\_move, Killed) = 0.5$

$T(Pipe, slow\_jump, Right\_side) = 0.25$
$T(Pipe, slow\_jump, Pipe) = 0.25$
$T(Pipe, slow\_jump, Killed) = 0.5$
$T(Pipe, slow\_move, Right\_side) = 0.25$
$T(Pipe, slow\_move, Pipe) = 0.25$
$T(Pipe, slow\_move, Killed) = 0.5$

$T(Right\_side, fast\_jump, Goal) = 1$
$T(Right\_side, fast\_move, Goal) = 1$
$T(Right\_side, slow\_jump, Goal) = 0.8$
$T(Right\_side, slow\_jump, Right\_side) = 0.2$
$T(Right\_side, slow\_move, Goal) = 0.8$
$T(Right\_side, slow\_move, Right\_side) = 0.2$

حالت های Killed و Goal حالت های نهایی ما هستند و در این مساله بازی در این حالت ها با اتمام میرسد برای همین هیچ احتمال گذاری برای آنها نوشته نشده است. هر چند میتوان گفت برای هر عملی در همان حالت باقی میمانند (با احتمال ۱و پاداش صفر)

پاداش های این مساله به این صورت تعریف میشوند:
پاداش مردن ماریو -۱۰۰ بوده و پاداش رسیدن به هدف +۱۰۰ است. بقیه حالات پاداشی ندارند.
در نهایت مدل MDP ما به صورت زیر خواهد بود:



سوال ۲)

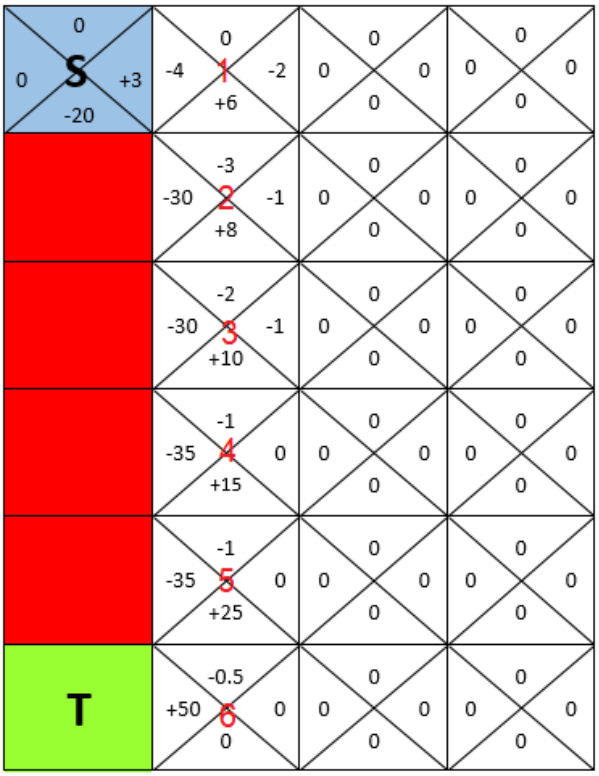
با استفاده از رابطه زیر برای هر اپیزود تک تک مراحل را اجرا کرده و مقادیر Q را بروز رسانی میکنیم.

$$New\ Q(s,a) = Q(s,a) + \alpha\,[R(s,a) + \gamma\ maxQ'(s',a') - Q(s,a)]$$

- New Q Value for that state and the action
- Learning Rate
- Reward for taking that action at that state
- Current Q Values
- Maximum expected future reward given the new state (s') and all possible actions at that new state.
- Discount Rate

برای راحتی کار روی حالت ها یک اسم میگذاریم:

1 2 3 4 5 6 بر روی حالت ها با رنگ قرمز نوشته شده است.



ابتدا اپیزود ۱ را اعمال کرده و مقادیر را بروزرسانی میکنیم.

عمل اول: Right

$$New\ Q(S,Right) = Q(S,Right) + \alpha * [R(S,Right) + \gamma * \max a\ Q\ (S,a) - Q(S,Right)]$$

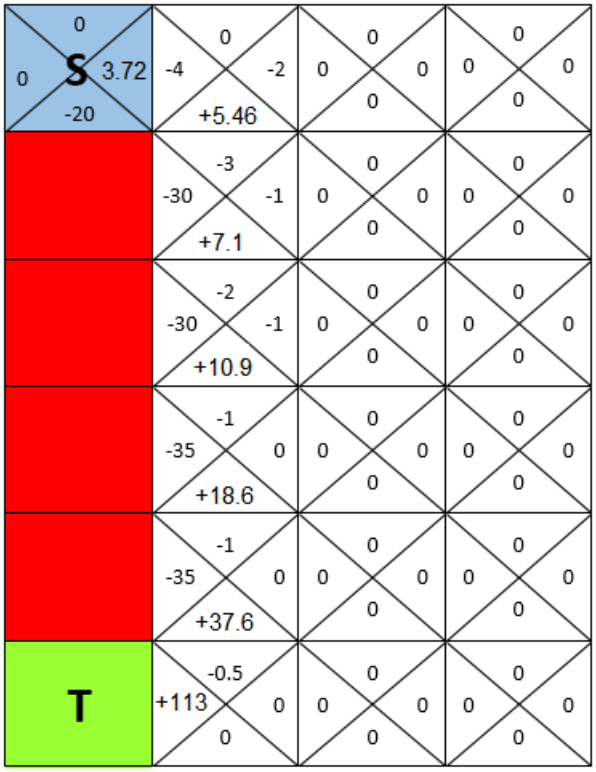
با جای گذاری مقادیر موجود مقدار جدید را محاسبه کرده و بروز رسانی میکنیم.

$$New\ Q(S, Right) = 3 + (0.9) * [-1 + (0.8) * 6 - 3] = 3.72$$

برای گام های بعدی در همان اپیزود داریم:

$$New\ Q(1, Down) = 6 + (0.9) * [-1 + (0.8) * 8 - 6] = 5.46$$
$$New\ Q(2, Down) = 8 + (0.9) * [-1 + (0.8) * 10 - 8] = 7.1$$
$$New\ Q(3, Down) = 10 + (0.9) * [-1 + (0.8) * 15 - 10] = 10.9$$
$$New\ Q(4, Down) = 15 + (0.9) * [-1 + (0.8) * 25 - 15] = 18.6$$
$$New\ Q(5, Down) = 25 + (0.9) * [-1 + (0.8) * 50 - 25] = 37.6$$
$$New\ Q(6, Left) = 50 + (0.9) * [120 + (0.8) * 0 - 50] = 113$$

پس از بروز رسانی های این اپیزود مقادیر بروز شده در محیط به صورت زیر خواهند بود.



حال اپیزود دو را اجرا کرده و مقادیر را بروز رسانی میکنیم.

$$New\ Q(S, Right) = 3.72 + (0.9) * [-1 + (0.8) * 5.46 - 3.72] = 3.4032$$
$$New\ Q(1, Down) = 5.46 + (0.9) * [-1 + (0.8) * 7.1 - 5.46] = 4.758$$
$$New\ Q(2, Down) = 7.1 + (0.9) * [-1 + (0.8) * 10.9 - 7.1] = 7.658$$
$$New\ Q(3, Left) = -30 + (0.9) * [-90 + (0.8) * 0 - (-30)] = -84$$

پس از بروز رسانی های این اپیزود مقادیر بروز شده نهایی در محیط به صورت زیر خواهند بود. (به دلیل جا شدن در شکل به صورت تقریبی با دو رقم اعشار نوشته شده اند.)

0 0 S +3.40 -20	0 -4 -2 +4.75	0 0 0 0	0 0 0 0	0 0 0 0
	-3 -30 -1 +7.65	0 0 0 0	0 0 0 0	0 0 0 0
	-2 -84 -1 +10.9	0 0 0 0	0 0 0 0	0 0 0 0
	-1 -35 0 +18.6	0 0 0 0	0 0 0 0	0 0 0 0
	-1 -35 0 +37.6	0 0 0 0	0 0 0 0	0 0 0 0
T	-0.5 +113 0 0	0 0 0 0	0 0 0 0	0 0 0 0

منابع:

https://stackoverflow.com/questions/58473521/how-do-i-calculate-maxq-in-q-learning

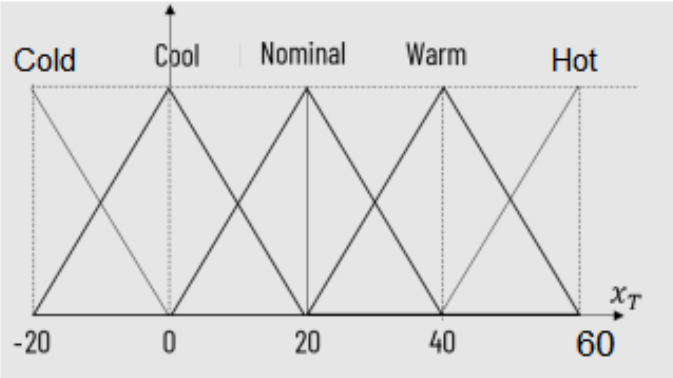
---

سوال ۳) حذف شده است.

سوال ۴)

ابتدا ورودی ها را به صورت فازی در آورده و برای آن نمودار درجه عضویت میکشیم.

برای ورودی دما (T) داریم:



همانطور که گفته شده است ۵ حالت فازی مختلف در نظر گرفته و به صورت مشهود در نمودار تقسیم بندی میکنیم.

برای یک مجموعه به طور مثال ضابطه را استخراج میکنیم و برای بقیه قسمت ها هم به همین ترتیب (به راحتی و با نسبت و تناسب مقدار درجه عضویت قابل محاسبه است.)

$$\mu_{Nominal} = \begin{cases} \frac{x_T}{20} & 0 \le x_T \le 20 \\ \frac{40 - x_T}{20} & 20 \le x_T \le 40 \end{cases}$$

به همین ترتیب برای بقیه زیر مجموعه ها هم ضابطه را محاسبه میکنیم.

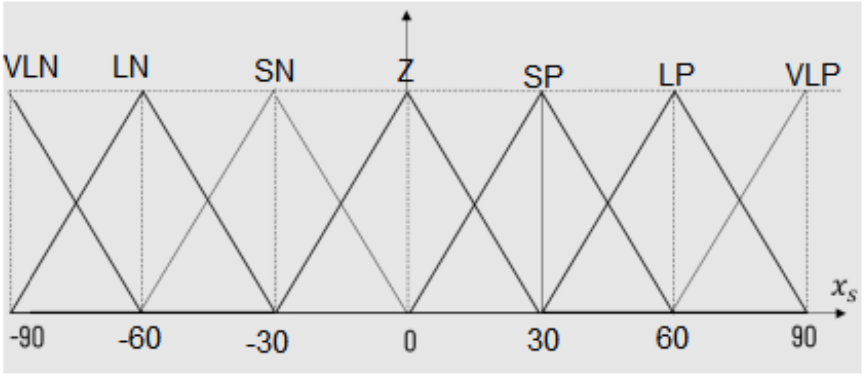
برای ورودی فشار (P) داریم:

همانطور که گفته شده است ۵ حالت فازی (VL خیلی کم – L کم – M متوسط – H بالا – VH بسیار بالا) مختلف در نظر گرفته و به صورت مشهود در نمودار تقسیم بندی میکنیم.
برای یک مجموعه به طور مثال ضابطه را استخراج میکنیم و برای بقیه قسمت ها هم به همین ترتیب (به راحتی و با نسبت و تناسب مقدار درجه عضویت قابل محاسبه است.)

$$\mu_L = \begin{cases} \frac{x_P}{25} & 0 \le x_P \le 25 \\ \frac{50 - x_P}{25} & 25 \le x_P \le 50 \end{cases}$$

به همین ترتیب برای بقیه زیر مجموعه ها هم ضابطه را محاسبه میکنیم.

برای خروجی هم به صورت زیر نمودار درجه عضویت را تعریف میکنیم:



همانطور که گفته شده است ۷ حالت فازی (VLN منفی بسیار بزرگ – LN منفی بزرگ – SN منفی کوچک – Zصفر - SP مثبت کوچک – LP مثبت بزرگ – VLP مثبت بسیار بزرگ) مختلف در نظر گرفته و به صورت مشهود در نمودار تقسیم بندی میکنیم.

برای یک مجموعه به طور مثال ضابطه را استخراج میکنیم و برای بقیه قسمت ها هم به همین ترتیب (به راحتی و با نسبت و تناسب مقدار درجه عضویت قابل محاسبه است.)

$$\mu_{SP} = \begin{cases} \frac{x_S}{30} & 0 \le x_S \le 30 \\ \frac{60 - x_S}{30} & 30 \le x_S \le 60 \end{cases}$$

به همین ترتیب برای بقیه زیر مجموعه ها هم ضابطه را محاسبه میکنیم.

در این مرحله جدول قوانین را ایجاد میکنیم. (توسط متخصص)

T \ P	VL	L	M	H	VH
Cold	VLP	VLP	Z	VLN	VLN
Cool	VLP	LP	Z	LN	VLN
Nominal	VLP	LP	Z	LN	VLN
Warm	LP	LP	Z	LN	LN
Hot	LP	SP	VLN	SN	LN

حال ورودی ها را حساب کرده و تبدیل به حالت فازی میکنیم تا از روی قوانین خروجی فازی را بدست آوریم:

منظور از ورودی دما ۷۰ درصد یعنی:

$$\left(\frac{70}{100} * 80\right) + (-20) = 36$$

که در دو حالت Warm و Nominal جای میگیرد. برای این حالت μ ها را محاسبه میکنیم.

$$\mu_{Warm} = \frac{16}{20} = 0.8, \qquad \mu_{Nominal} = \frac{20 - 16}{20} = 0.2$$

منظور از ورودی فشار ۳۰ درصد یعنی:

$$\left(\frac{30}{100} * 100\right) + 0 = 30$$

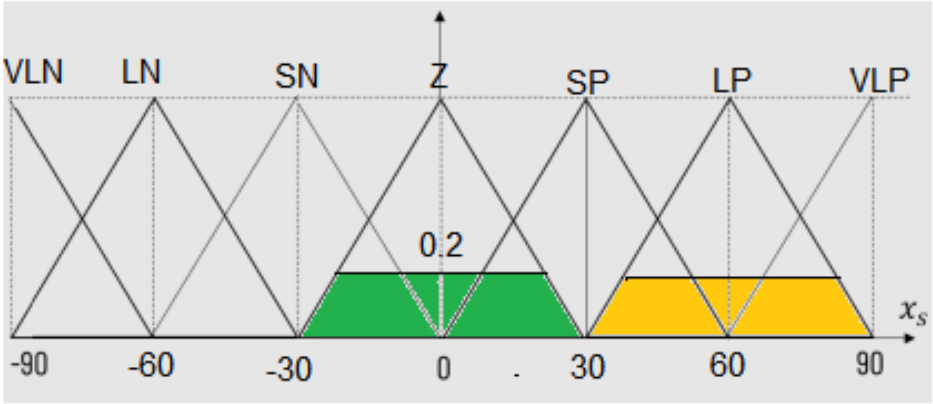
که در دو حالت M و L جای میگیرد. برای این حالت μ ها را محاسبه میکنیم.

$$\mu_M = \frac{30 - 25}{25} = 0.2, \qquad \mu_L = \frac{50 - 30}{25} = 0.8$$

حال چهار حالت موجود را بر اساس قوانین معلوم کرده و خروجی را مشخص میکنیم.
برای ترکیب درجه عضویت ها از قاعده Mamdani استفاده کرده و minimum میگیریم. (برای and)

قانون ۱:
دما Nominal با درجه 0.2 و فشار L با درجه 0.8 : خروجی LP با درجه 0.2
قانون ۲:
دما Nominal با درجه 0.2 و فشار M با درجه 0.2 : خروجی Z با درجه 0.2
قانون ۳:
دما Warm با درجه 0.8 و فشار L با درجه 0.8 : خروجی LP با درجه 0.2
قانون ۴:
دما Warm با درجه 0.8 و فشار M با درجه 0.2 : خروجی Z با درجه 0.2

حال باید نتایج را aggregate کنیم. با قاعده ی MaxMin جلو میرویم و در نهایت بر روی نمودار داریم:
(که ناحیه Z با درجه 0.2 و ناحیه LP هم با درجه 0.2 جزو جواب ما خواهند بود)



با روش میانگین وزن دار مرکز را محاسبه میکنیم که جواب نهایی ما خواهد بود.

$$\mu_{x1} = 0.2, \mu_{x2} = 0.2$$
$$\bar{x}_1 = 0, \bar{x}_2 = 60$$
$$x^* = \frac{\bar{x}_1\mu_{x1} + \bar{x}_2\mu_{x2}}{\mu_{x1} + \mu_{x2}} = 30$$

پس کنترلر ما در نهایت برای حالت گفته شده مقدار ۳۰ را برای خروجی پیشنهاد میدهد.

منابع:

https://codecrucks.com/designing-fuzzy-controller-step-by-step-guide/
https://www.youtube.com/watch?v=byjJNoJMQ48

سوال ۵) در نوت بوک پیوست شده پیاده سازی انجام شده است.
