# Symbolic Network: Generalized Neural Policies for Relational MDPs Supplementary Paper

**Sankalp Garg** [1] **Aniket Bajpai** [1] **Mausam** [1]

## A. Domain Description

We describe the details of the domains presented in the IPPC 2011 and IPPC 2014. The statistics for state fluents ($\mathcal{F}$), non-fluents ($\mathcal{NF}$) and Action ($\mathcal{A}$) for all the domains are show in the Table 1 and Table 2. UP represent $\mathcal{F}, \mathcal{NF}$ and $\mathcal{A}$ without parameters, Unary represents $\mathcal{F}, \mathcal{NF}$ and $\mathcal{A}$ with a single parameter and multiple represents $\mathcal{F}, \mathcal{NF}$ and $\mathcal{A}$ with more than one parameter. Table 3 lists the instance specific number of objects, state variables and action variables for the domain. The domains $1, 2, 3$ are used for training, $4$ for validation and $5, 6, 7, 8, 9, 10$ for testing.

### Academic Advising

The academic advising domain represents a student at a university trying to complete his/her degree. Some courses are required to be completed to obtain the final degree. Each course is either a basic course or may have prerequisites. The probability of passing a course depends on the number of prerequisites completed (a fixed probability if no prerequisite). The goal is to complete the degree as soon as possible.

### Crossing Traffic

Crossing Traffic is represents a robot in a grid, with obstacles at a random grid cell at any time. The obstacles (car) start at any cell randomly and move left. The robot aims to plan its path from the starting grid cell to the goal cell while avoiding obstacles.

### Game of Life

Game of Life domain is represented as a grid where each cell can either be dead or alive. The goal is to keep as many cells alive as possible. The probability of cell death depends on the number of neighbors alive at a particular time, which is non-linear in the number of neighbors alive.

Table 1: The statistics related to the domains listing the number of UP (Un-Paramataried), Unary and Muiltiple Action ($\mathcal{A}$) for each domain.

| Domain | UP-$\mathcal{A}$ | Unary-$\mathcal{A}$ | Multiple-$\mathcal{A}$ |
|---|---|---|---|
| Academic Advising | 0 | 1 | 0 |
| Crossing Traffic | 4 | 0 | 0 |
| Game of Life | 0 | 0 | 1 |
| Navigation | 4 | 0 | 0 |
| Skill Teaching | 0 | 2 | 0 |
| Sysadmin | 0 | 1 | 0 |
| Tamarisk | 0 | 2 | 0 |
| Traffic | 0 | 1 | 0 |
| Wildfire | 0 | 0 | 2 |

### Navigation

Navigation represents a robot in a grid world where the aim is to reach a goal cell as quickly as possible. The probability of the robot dying in a particular cell is different, which is specified in the instance file.

### Skill Teaching

Skill Teaching domain represents a teacher trying to teach a skill to students. Each student has a mastery level in a particular skill. Some skills have pre-conditions, which increase the probability of learning a particular skill. The skill is taught using either hints or multiple-choice questions. The goal is to answer as many questions as possible by the student by learning the required skill.

### Sysadmin

Sysadmin domain represents computers connected in a network. The probability of a computer shutting down on its own depends on the number of turned-on neighboring computers. The agent can either turn on a computer or leave it as it is. The goal is to maximize the number of computers at a particular time.

[1]Indian Institute of Technology Delhi. Correspondence to: Sankalp Garg <sankalp2621998@gmail.com>, Aniket Bajapi <quantum.computing96@gmail.com>, Mausam <mausam@cse.iitd.ac.in>.

*Proceedings of the $37^{th}$ International Conference on Machine Learning*, Online, PMLR 119, 2020. 

Table 2: The statistics related to the domains listing the number of UP (Un-Paramataried), Unary and Multiple State Fluents ($\mathcal{F}$) and Non-Fluents ($\mathcal{NF}$) for each domain.

| Domain | UP-$\mathcal{F}$ | UP-$\mathcal{NF}$ | Unary-$\mathcal{F}$ | Unary-$\mathcal{NF}$ | Multiple-$\mathcal{F}$ | Multiple-$\mathcal{NF}$ |
|---|---|---|---|---|---|---|
| Academic Advising | 0 | 1 | 2 | 5 | 0 | 1 |
| Crossing Traffic | 0 | 1 | 0 | 4 | 2 | 5 |
| Game of Life | 0 | 0 | 0 | 0 | 1 | 2 |
| Navigation | 0 | 0 | 0 | 4 | 1 | 6 |
| Skill Teaching | 0 | 0 | 6 | 7 | 0 | 1 |
| Sysadmin | 0 | 2 | 1 | 0 | 0 | 1 |
| Tamarisk | 0 | 17 | 2 | 0 | 0 | 2 |
| Traffic | 0 | 0 | 3 | 3 | 0 | 3 |
| Wildfire | 0 | 4 | 0 | 0 | 2 | 2 |

**Tamarisk**

Tamarisk domain represents invasive species of plants (Tamarisk) trying to take over native plant species. The plants spread in any direction and try to destroy the native plant species. The agent can either eradicate Tamarisk in a cell or restore the native plant species, each having a different reward. The goal is to minimize the cost of eradication and restoration of the native plant species.

**Traffic**

Traffic domain models the traffic on the road with roads connecting at various intersections. Each road intersection has two traffic light signals combinations of which yield different traffic movement. The agent aims to control the traffic signal (only on the forward sequence) to control the traffic.

**Wildfire**

The wildfire domain represents a forest catching fire. The direction of fire spreading depends on the direction of the wind and also the type of fuel at that point (e.g., grass or wood, etc.). The agent can either choose to put down the fire or cut off the fuel even before the fire happens. The goal is to prevent as many cells as possible, and more reward is provided to protect high priority cells.

## B. Variation of $\alpha_{\text{SymNet}}(0)$ with neighbourhood

To inspect the importance of the neighborhood information in learning a generalized policy for the domains, we perform the study of the neighborhood parameter variation. In the Figure 1, we show the variation of $\alpha_{\text{SymNet}}(0)$ with neighbourhood. From the Figure, we observe that message passing for the neighborhood of size 1 yields the best results for most domains, and hence we reported the results with neighborhood 1 in the main paper. In general, we observe that the value of $\alpha_{\text{SymNet}}(0)$ first increases and then decreases.

For most instances, the $\alpha_{\text{SymNet}}(0)$ is less for neighborhood 0 compared to neighborhood 1, showing that the information regarding the neighbors is necessary for learning a better policy. For example, in domain academic advising, the neighborhood 1 aggregates information about the prerequisites for the courses and then prioritizes the courses to take. A similar trend is observed in domain skill teaching, where the information about the pre-condition for the skill plays an important role in learning the skills. For some domains like navigation, neighborhood information is absolutely critical for planning the next move which can be observed from very low values of $\alpha_{\text{SymNet}}(0)$ from Figure 1(d). Other domains like wildfire are not affected a lot by neighborhood a lot. This is because the margin between the minimum and maximum rewards is large, and the generalized policy outputs rewards close to the maximum value, which decreases the variation in the value of $\alpha_{\text{SymNet}}(0)$. As we increase the value of neighborhood to 2 and 3, the value of $\alpha_{\text{SymNet}}(0)$ tends to fall down for most instances. We hypothesize that the agent overfits to instance-specific policies for the instances it is trained on and hence fails to generalize.

Table 3: The statistics related to the domain instances listing the number of Objects, State Variables and Action Variables for all the instances of the domains. Domain $1, 2, 3$ are used for training, $4$ for validation and $5, 6, 7, 8, 9, 10$ for testing.

| Domain | #Objects | #State Vars | #Action Vars | Domain | #Objects | #State Vars | #Action Vars |
|---|---|---|---|---|---|---|---|
| AA 1 | 10 | 20 | 11 | ST 1 | 2 | 12 | 5 |
| AA 2 | 10 | 20 | 11 | ST 2 | 2 | 12 | 5 |
| AA 3 | 15 | 30 | 16 | ST 3 | 4 | 24 | 9 |
| AA 4 | 15 | 30 | 16 | ST 4 | 4 | 24 | 9 |
| AA 5 | 20 | 40 | 21 | ST 5 | 6 | 36 | 13 |
| AA 6 | 20 | 40 | 21 | ST 6 | 6 | 36 | 13 |
| AA 7 | 25 | 50 | 26 | ST 7 | 7 | 42 | 15 |
| AA 8 | 25 | 50 | 26 | ST 8 | 7 | 42 | 15 |
| AA 9 | 30 | 60 | 31 | ST 9 | 8 | 48 | 17 |
| AA 10 | 30 | 60 | 31 | ST 10 | 8 | 48 | 17 |
| CT 1 | 9 | 12 | 5 | Sys 1 | 10 | 10 | 11 |
| CT 2 | 9 | 12 | 5 | Sys 2 | 10 | 10 | 11 |
| CT 3 | 16 | 24 | 5 | Sys 3 | 20 | 20 | 21 |
| CT 4 | 16 | 24 | 5 | Sys 4 | 20 | 20 | 21 |
| CT 5 | 25 | 40 | 5 | Sys 5 | 30 | 30 | 31 |
| CT 6 | 25 | 40 | 5 | Sys 6 | 30 | 30 | 31 |
| CT 7 | 36 | 60 | 5 | Sys 7 | 40 | 40 | 41 |
| CT 8 | 36 | 60 | 5 | Sys 8 | 40 | 40 | 41 |
| CT 9 | 49 | 84 | 5 | Sys 9 | 50 | 50 | 51 |
| CT 10 | 49 | 84 | 5 | Sys 10 | 50 | 50 | 51 |
| GOL 1 | 9 | 9 | 10 | Tam 1 | 12 | 16 | 9 |
| GOL 2 | 9 | 9 | 10 | Tam 2 | 16 | 24 | 9 |
| GOL 3 | 9 | 9 | 10 | Tam 3 | 15 | 20 | 11 |
| GOL 4 | 16 | 16 | 17 | Tam 4 | 20 | 30 | 11 |
| GOL 5 | 16 | 16 | 17 | Tam 5 | 18 | 24 | 13 |
| GOL 6 | 16 | 16 | 17 | Tam 6 | 24 | 36 | 13 |
| GOL 7 | 25 | 25 | 26 | Tam 7 | 21 | 28 | 15 |
| GOL 8 | 25 | 25 | 26 | Tam 8 | 28 | 42 | 15 |
| GOL 9 | 25 | 25 | 26 | Tam 9 | 24 | 32 | 17 |
| GOL 10 | 30 | 30 | 31 | Tam 10 | 32 | 48 | 17 |
| Nav 1 | 12 | 12 | 5 | Tra 1 | 28 | 32 | 5 |
| Nav 2 | 15 | 15 | 5 | Tra 2 | 28 | 32 | 5 |
| Nav 3 | 20 | 20 | 5 | Tra 3 | 40 | 44 | 5 |
| Nav 4 | 30 | 30 | 5 | Tra 4 | 40 | 44 | 5 |
| Nav 5 | 30 | 30 | 5 | Tra 5 | 52 | 56 | 5 |
| Nav 6 | 40 | 40 | 5 | Tra 6 | 52 | 56 | 5 |
| Nav 7 | 50 | 50 | 5 | Tra 7 | 64 | 68 | 5 |
| Nav 8 | 60 | 60 | 5 | Tra 8 | 64 | 68 | 5 |
| Nav 9 | 80 | 80 | 5 | Tra 9 | 76 | 80 | 5 |
| Nav 10 | 100 | 100 | 5 | Tra 10 | 76 | 80 | 5 |
| Wild 1 | 9 | 18 | 19 | Wild 6 | 25 | 50 | 51 |
| Wild 2 | 9 | 18 | 19 | Wild 7 | 30 | 60 | 61 |
| Wild 3 | 16 | 32 | 33 | Wild 8 | 30 | 60 | 61 |
| Wild 4 | 16 | 32 | 33 | Wild 9 | 36 | 72 | 73 |
| Wild 5 | 25 | 50 | 51 | Wild 10 | 36 | 72 | 73 |