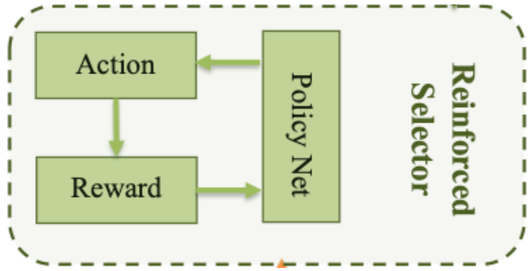
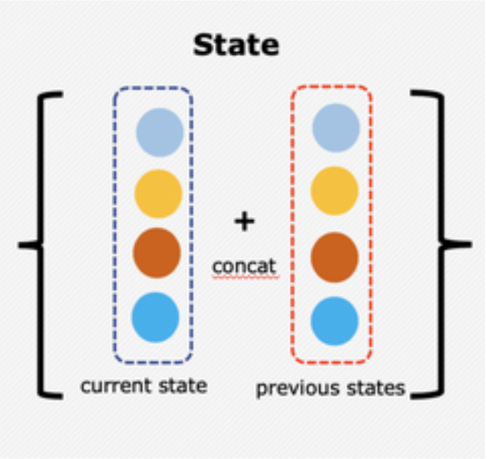
Methodology

Data Selection via Reinforcement Learning - *State*



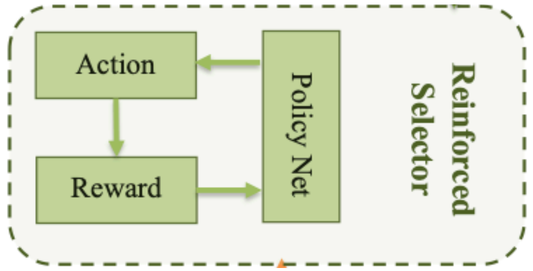
- The current state vector contains four elements:

  1) output probability from the annotator (quality) ▪
  2) output probability from fake news detector (quality) ▪
  3) maximum of cosine similarity between the current sample and the chosen samples (diversity) ▪
  4) weak label of the current sample (balance the distribution of classes) ▪



Methodology

Data Selection via Reinforcement Learning - *Action*



- The action value $a_i^{(k)}$ for every sample is 1 or 0.
  - 1: the action to *retain* the sample
  - 0: the action to *remove* the sample
- To determine the action, train a policy network includes two fully connected layers with corresponding activation functions, denote as $P\left(\cdot;\theta_s\right)$, θ_s: the parameters
  - $P\left(s_i^{(k)};\theta_s\right)=\delta\left(\mathbf{w}_{s2}\cdot\mathrm{ReLU}\left(\mathbf{w}_{s1}\cdot s_i^{(k)}\right)\right)$
  - $\mathbf{w}_{s1},\mathbf{w}_{s2}$: weights of fully-connected layer, δ: sigmoid activation function