An Ensemble Learning Method based on Q learning

Kharazmi University

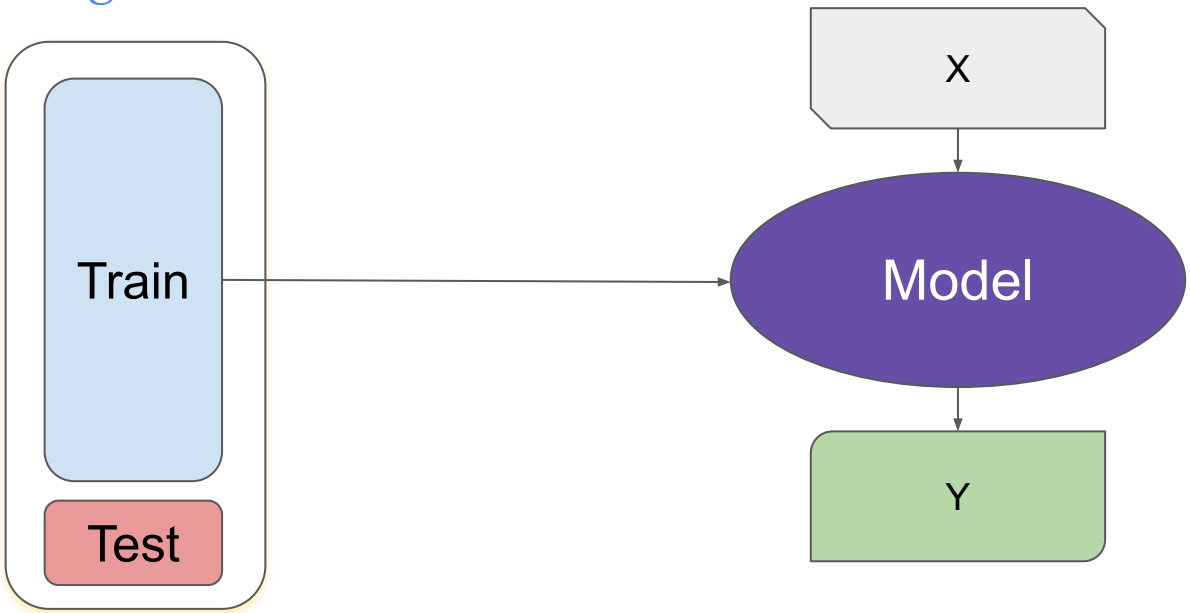
AmirMohammad Hosseini Nasab
AmirMahdi Mohseni

Contents

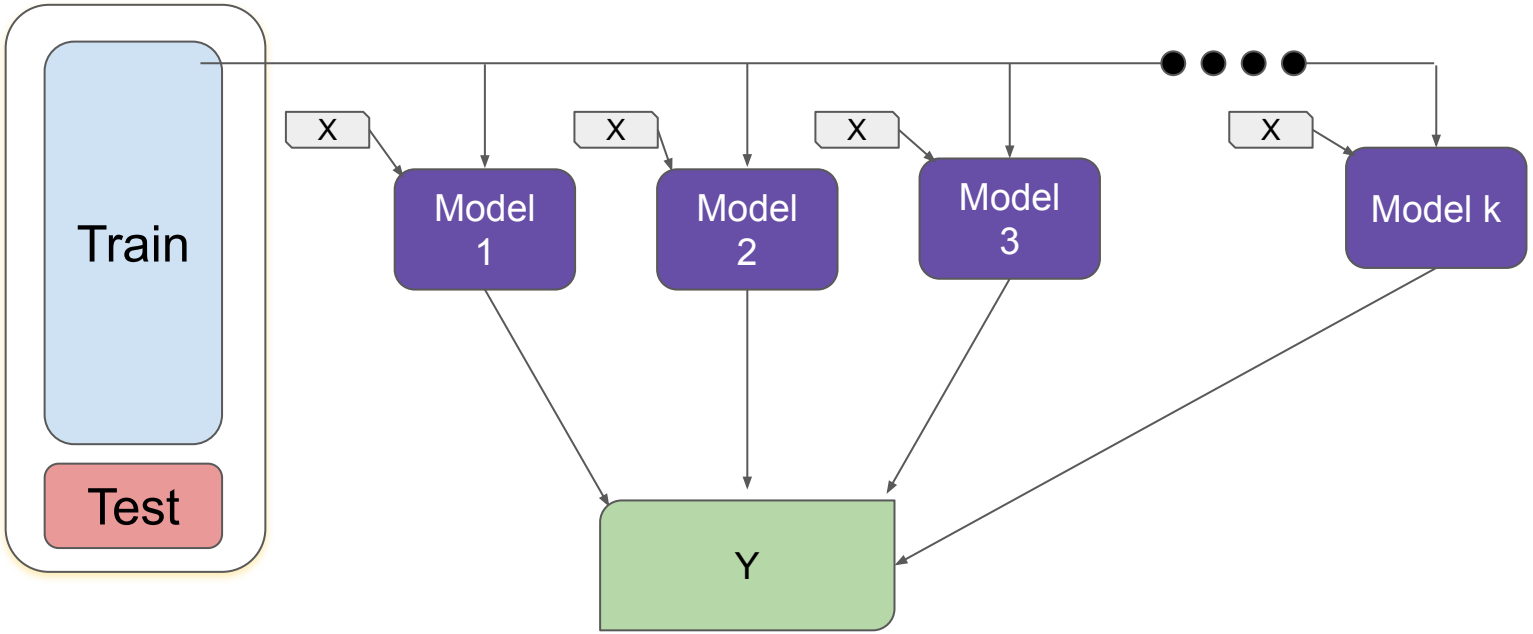
- Ensemble Learning
- Q learning
- Sentiment Analysis
- Proposed Method

Ensemble Learning

A single model learner



Ensemble Learner



Ensemble Learner

- There are three main categories:
  - Bagging
  - Boosting
  - Random Subspace

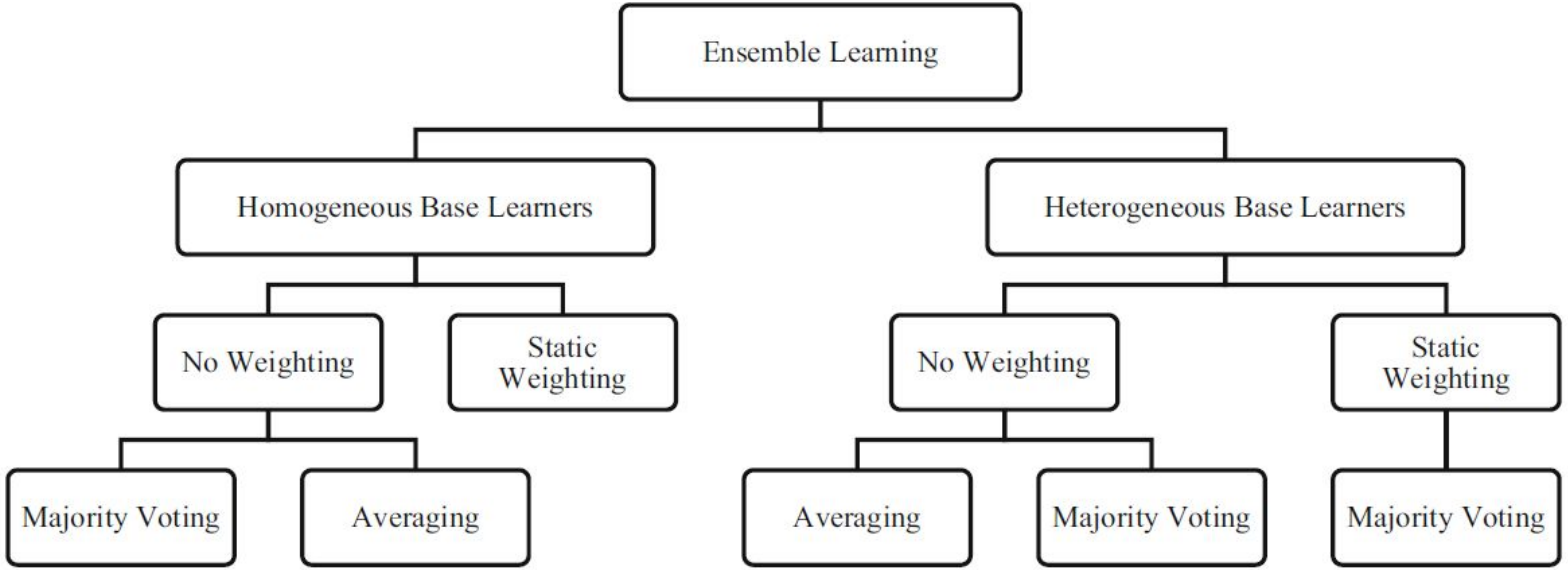
- Categorized by base learners:
  - Homogeneous base learners
  - Heterogeneous base learners

Ensemble Learning

Output Y:

- Averaging
- Weighted voting
- Majority voting
- Stacking
- Bagging

Ensemble Learning



Q learning

A reinforcement learning technique

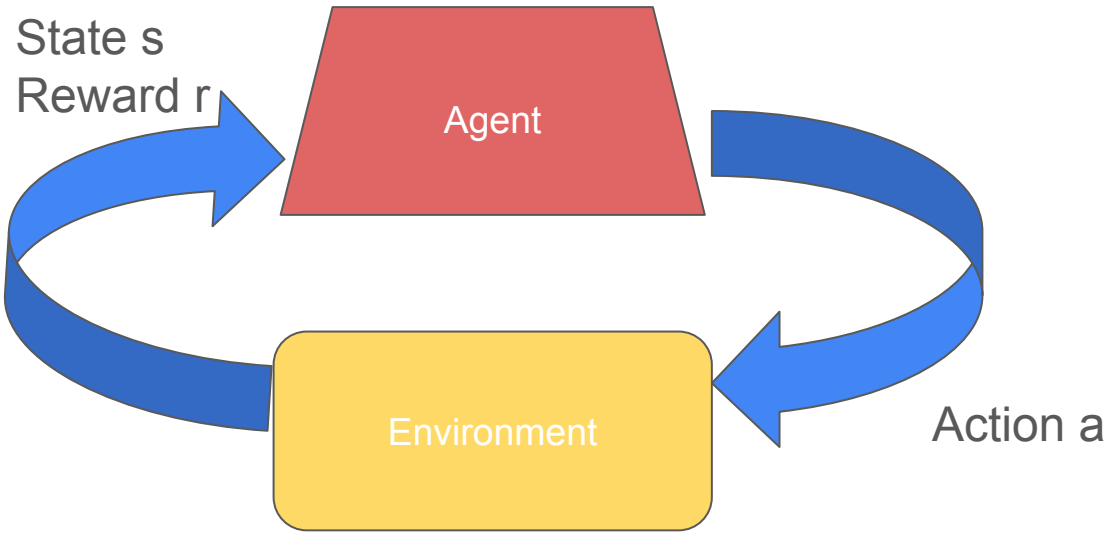
how an intelligent agent should take actions in a dynamic environment in order to maximize a reward signal

Determines a specific policy for taking actions in different situations

No models of the environment are required!

Agent and an Environment

An agent interacts with the environment

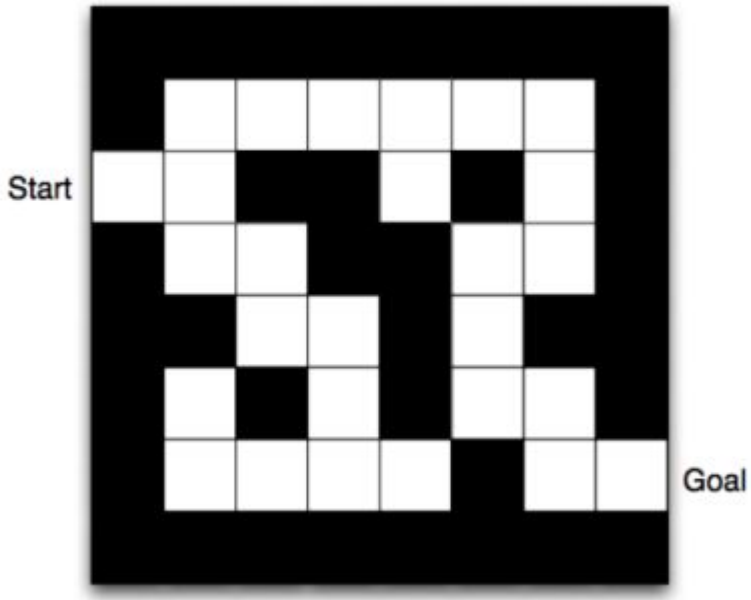


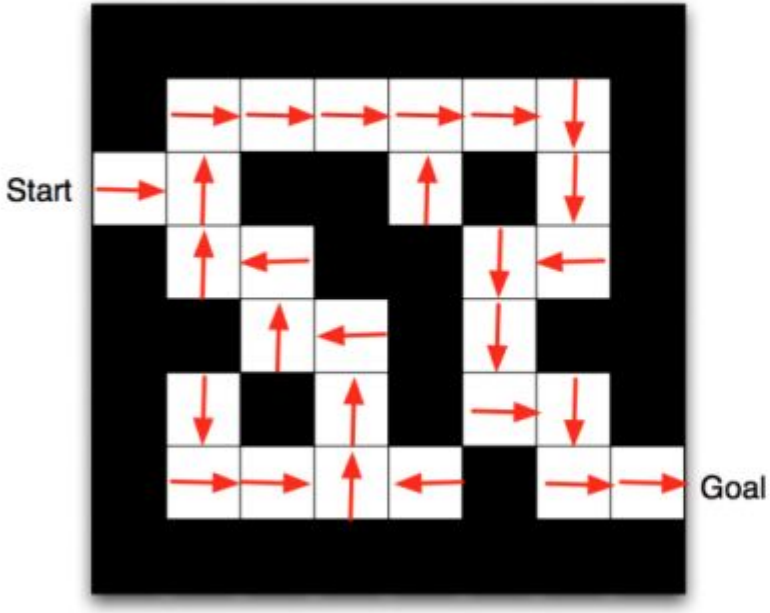
State : Location

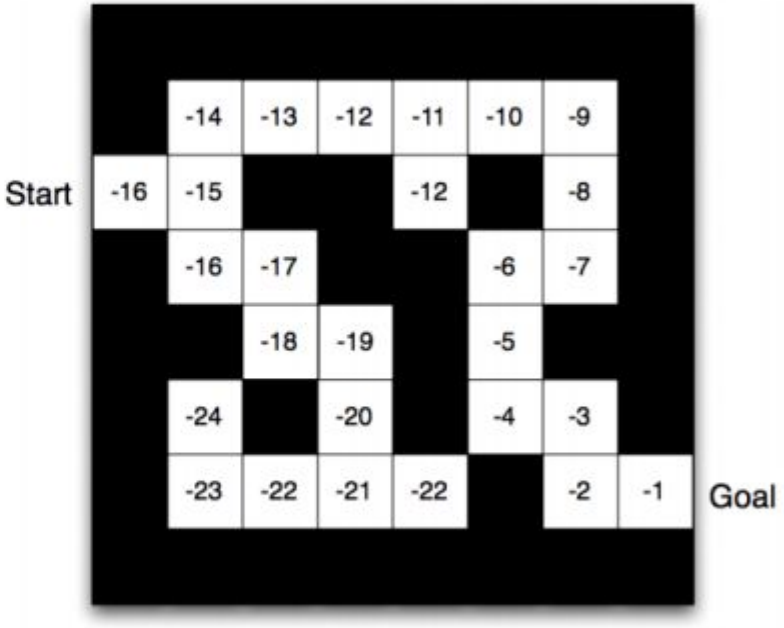
Actions : U, D, L, R

Reward : -1 per time step

Undiscounted




Start
Goal


Start
-14 -13 -12 -11 -10 -9
-16 -15 -12 -8
-16 -17 -6 -7
-18 -19 -5
-24 -20 -4 -3
-23 -22 -21 -22 -2 -1
Goal

Markov Decision Process (MDP)

Components of an MDP:

- Initial state distribution $p(s_0)$
- Transition distribution $p(s_{t+1} | s_t, a_t)$
- Reward function $r(s_t, a_t)$
- Discounted return
- $$G_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots$$
- Value Function
- $$V^{\pi}(s) = \mathbb{E}\left(\sum_{0}^{\infty} \gamma^i r_{t+i} \mid s = s_t \right)$$

Discount Factor

Determines whether future rewards are important to the agent

$$\gamma = 0 \rightarrow \textit{Only immediate rewards are considered}$$

$$\gamma = 1 \rightarrow \textit{future rewards are as important as immideate ones}$$

$$0 < \gamma < 1 \rightarrow \textit{determines the importance of future rewards}$$

Q learning

Can we find a value function to choose actions?

$$\arg\max_{\mathbf{a}} r(\mathbf{s}_t, \mathbf{a}) + \gamma \mathbb{E}_{p(\mathbf{s}_{t+1} \mid \mathbf{s}_t, \mathbf{a}_t)}[V^{\pi}(\mathbf{s}_{t+1})]$$

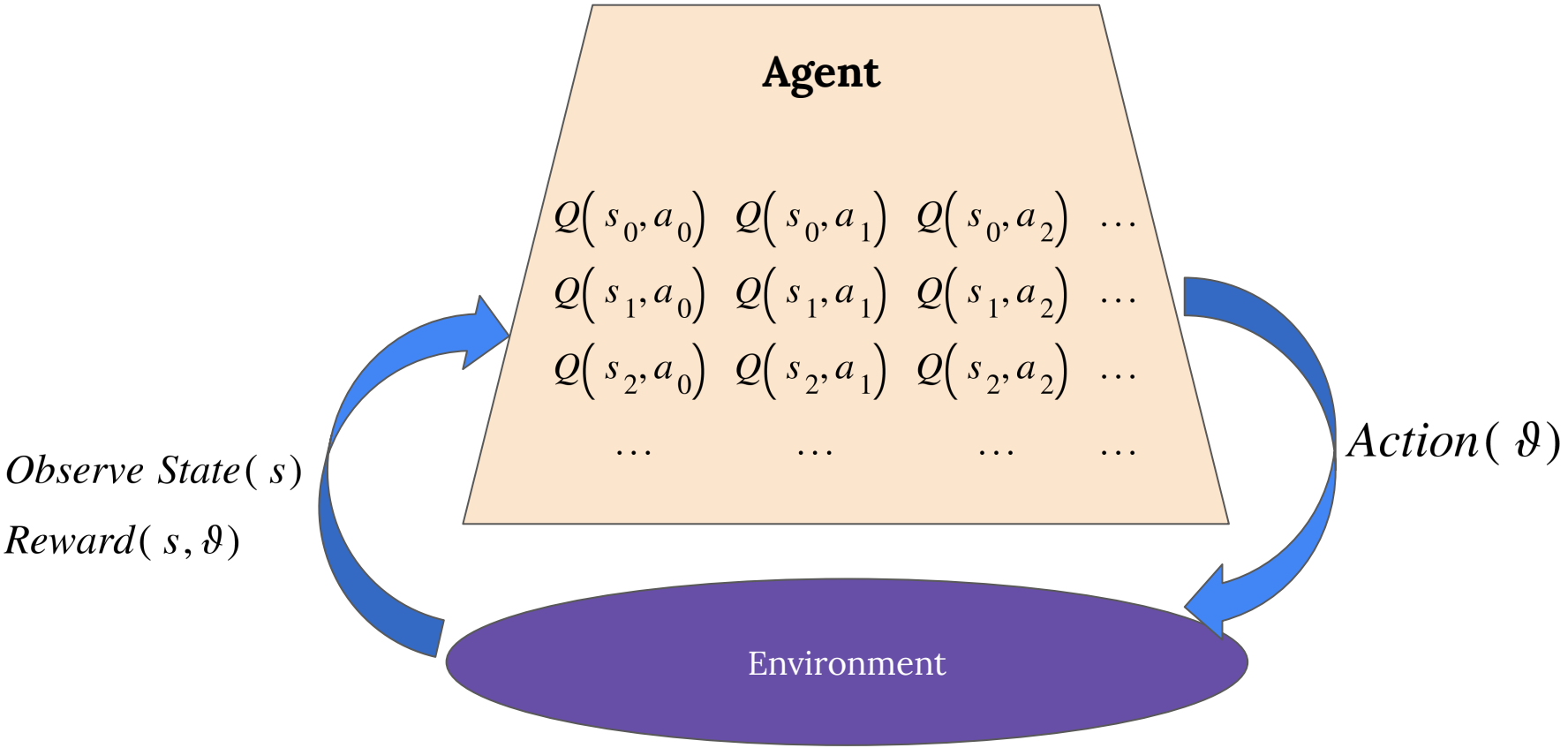
Instead we'll return the expected when taking action and then following the policy

$$Q^{\pi}(\mathbf{s}, \mathbf{a}) = \mathbb{E}[G_t \mid \mathbf{s}_t = \mathbf{s}, \mathbf{a}_t = \mathbf{a}]$$

$$V^{\pi}(\mathbf{s}) = \sum_{\mathbf{a}} \pi(\mathbf{a} \mid \mathbf{s}) Q^{\pi}(\mathbf{s}, \mathbf{a})$$

Maps each state to an action

Q learning Table



Bellman equation

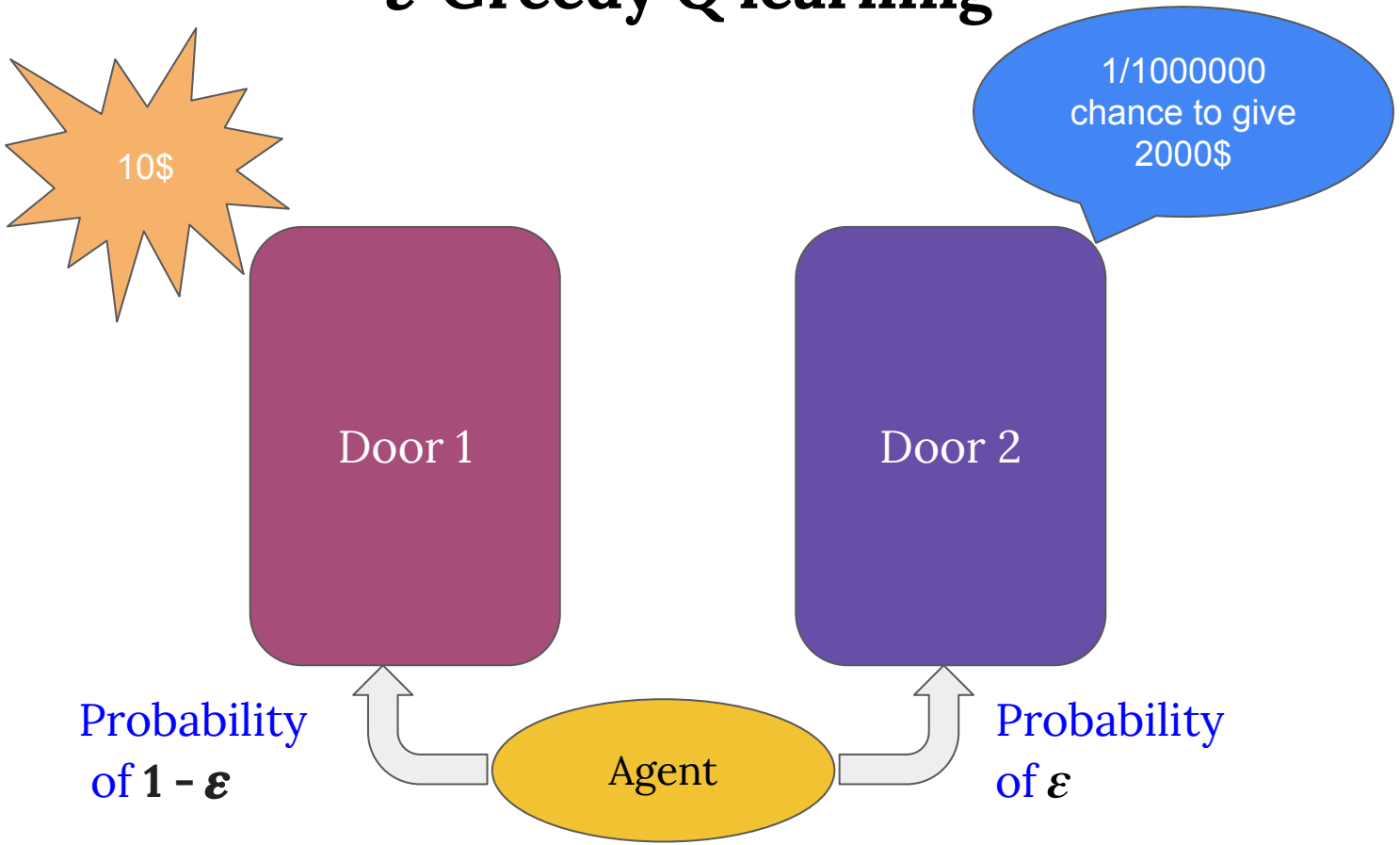
$$Q_{t+1}(s,a) = (1-\alpha)\,Q_t(s,a) + \alpha\left[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right]$$

Algorithm Q-Learning

1: **Initialize** $Q_t(s,a)$ // *arbitrarily*
2: **Repeat** (**for each** *episode*)
3: **Initialize** S randomly
4: **Repeat** (**for each** *step*)
5: **Select** an *action* a_i
6: **Execute** the *action* a
7: **Observe** r (s, a), S'
8: **Update** Q-Table according to **Eq. (1)**
9: $s \leftarrow s'$
10: **Until** S is the *terminate state*
11: **Until** some *stopping criteria* are reached

Q is convergent to the best rewards possible

ε-Greedy Q learning



Sentiment Analysis

Sentiment Analysis is a method for classifying text polarity into three levels of document, sentence or aspect. A data-driven process which involves various techniques such as NLP.

In examining a document or a sentence, it is assumed that only one sentiment is expressed.(Positive & Negative)

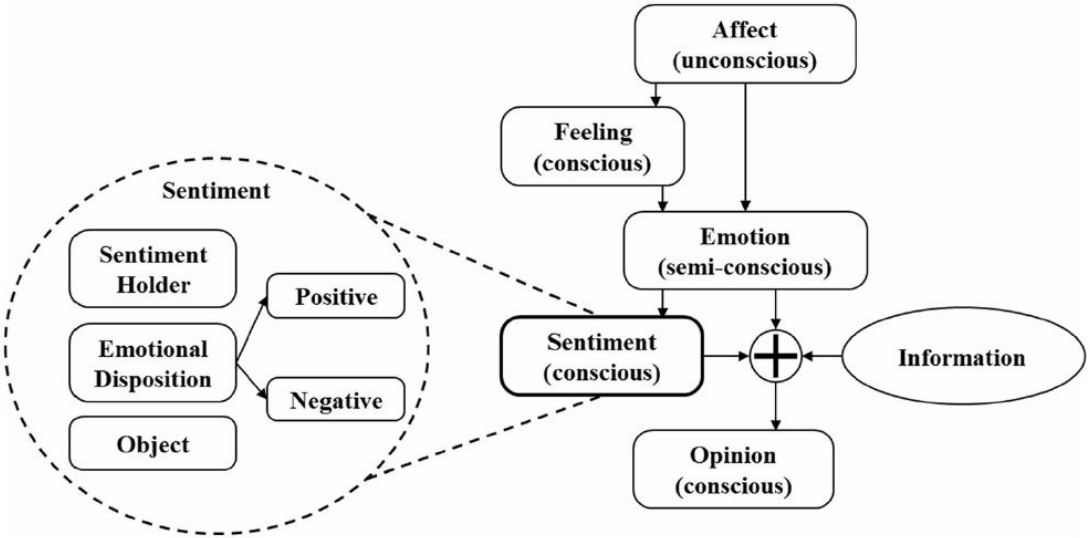
Assign a numerical score to a text document(1 & 0)

Notes

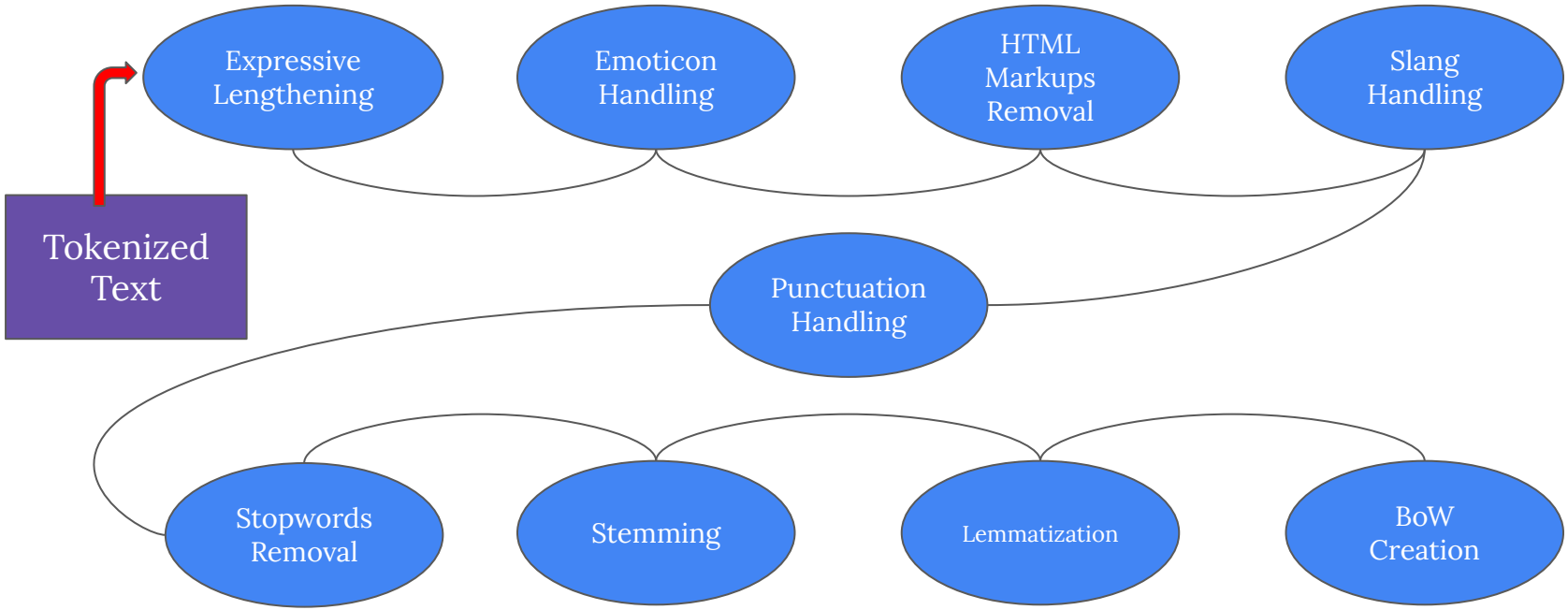
- Sentiment refers to the overall positive or negative tone of the text.
- Affect is the emotional state conveyed by some text
- Some algorithms use affect as a proxy for sentiment.
- Some algorithms infer feelings (which is the subjective experience i.e “happy”, “sad”, ”angry”) to describe events.
- Emotion refers to a complex psychological state.
- Opinion refers to the person’s subjective evaluation of a particular topic.
- Traditional sentiment analysis techniques involve fundamental challenges (Some texts could be sarcastic)
- Hence there’s a need for reinforcement techniques to make decisions based on problem space.

Method	Characteristics
Dictionary-based method	• Traditional classification method: Domain dependency • Aspect-based sentiment analysis: Requires fully organized data in specific domains (Wu et al. 2021), (Pham and Le 2018), (Song et al. 2021) • Unigram and N-gram sentiment analysis: dependency on the predefined domain of data • Hierarchical sentiment analysis: Hierarchical structure determination and managing communication between different levels and nodes
Reinforcement learning-based method	• No domain dependency (Beigi and Moattar 2021) • No need for basic knowledge about data • Ability to manage data with dynamic behavior

Sentiment Analysis Hierarchy



Text Processing



Preprocessing	Description	Example	
		Input	Output
Tokenization	Break a sentence into words, phrases, symbols, or other meaningful tokens	"This is my new car"	"This", "is", "my", "new", "car"
Expressive lengthening	Replaced word with the original form if written in the repetition of one of the letters	"Haaaaappy" "preeeeeetty"	"Happy", "pretty""
Emoticons handling	Replace emoticons with their meanings	:-), 8–0,:-*	"basic smiley", "oh my god", "kiss"
HTML Markups removal	Remove HTML markups	< p > , < / p > , < br > , < /br >	will remove these markups
Slangs handling	Replace slang with their original words	"wassup", "gr8", "2mrw"	"what's up?", "great", "tomorrow"
Punctuation handling	Punctuations and numbers are removed except apostrophes	" n't ", " 's", " 're"	"not", "is", "are"
Stopwords removal	Remove Stopwords to simplified text	"It was the best movie I have ever seen."	"best movie ever seen"
Stemming	Remove various prefixes and suffixes, to reduce the number of words	"'user", "users", "used", "using"	"use"
Lemmatization	Return the base or dictionary form of a word, which is known as the lemma	"looks", "was"	"look", "be"

BoW Creation

A Bow(Bag of Words) is a table in which the number of repetitions for each word is inserted.

	about	bird	heard	is	the	word	you
You heard about the bird.	1	1	1	0	1	0	1
The bird is the word	0	1	0	1	1	1	0
About bird bird bird	1	3	0	0	0	0	0

Proposed Method

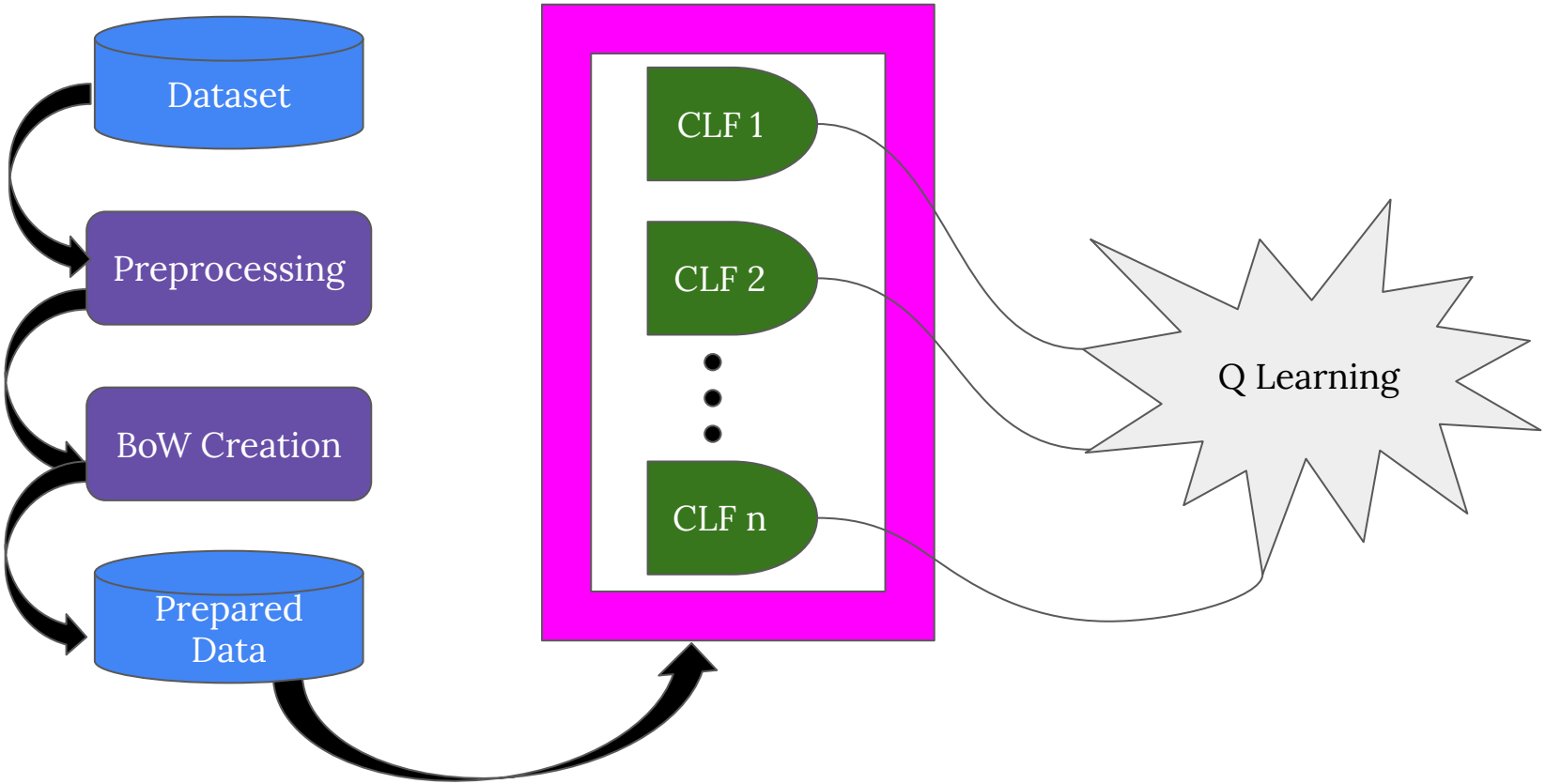
The set of actions that we can take at each step corresponds to choosing one of several base learners that provides the most achievable reward.

The choice of the base learner is based on Boltzmann's equation

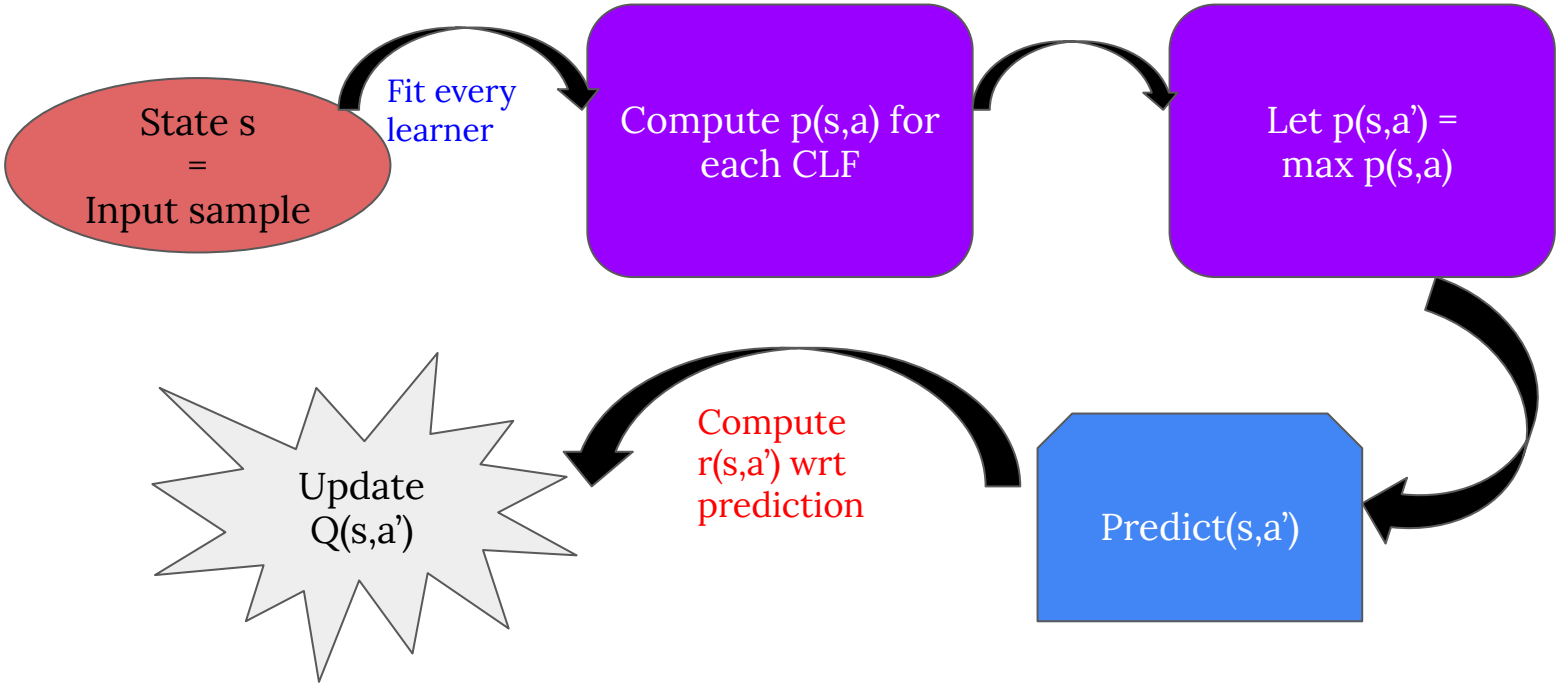
$$p(s,a) = \frac{e^{\frac{Q(s,a)}{\tau}}}{\sum_{a' \in A} e^{\frac{Q(s,a')}{\tau}}}$$

$$\tau = 0 : Greedy,\ \tau = \infty : Random,\ 0 < \tau < \infty : \varepsilon - greedy$$

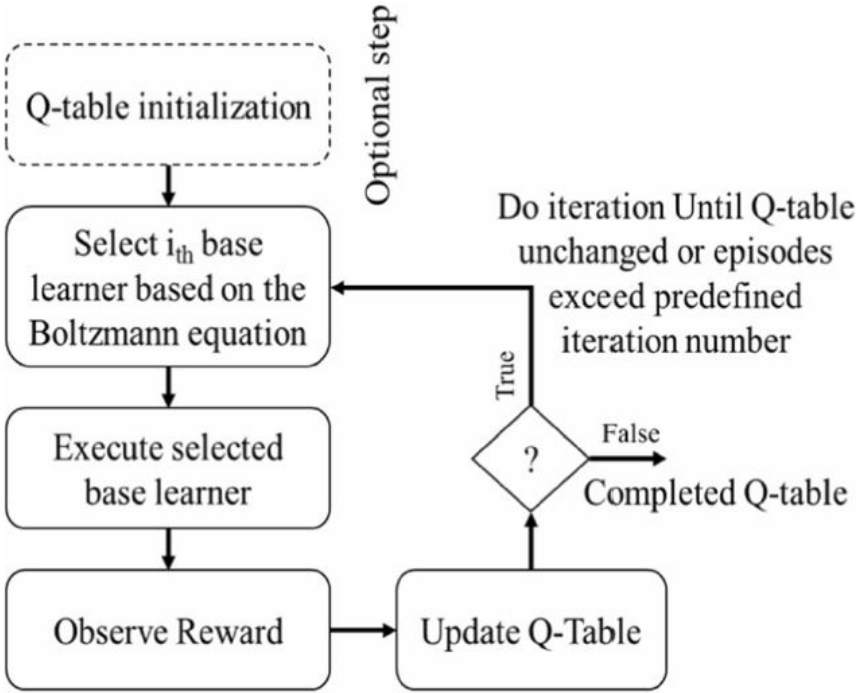
Proposed Method Diagram



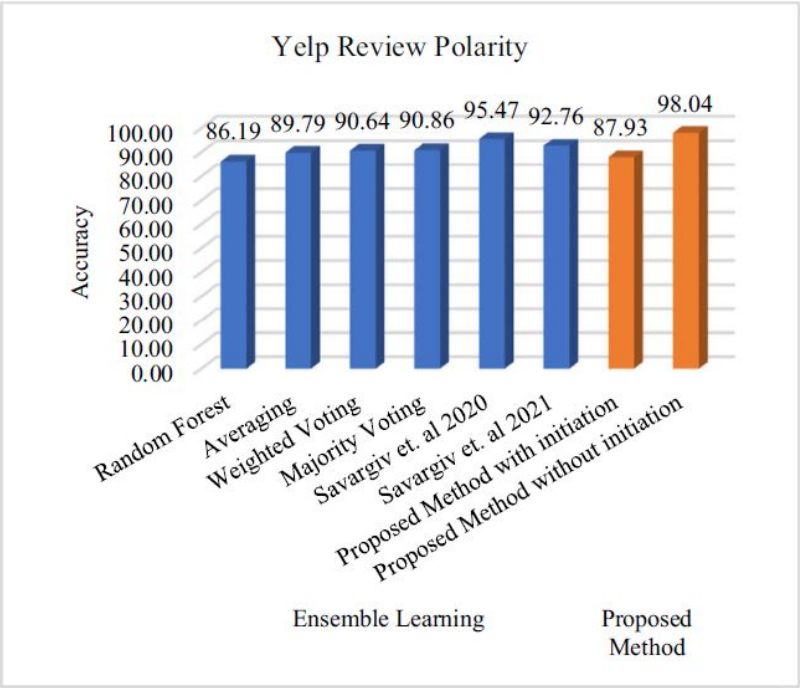
Q Learning Component



Overview


Optional step
Q-table initialization
Select i_{th} base learner based on the Boltzmann equation
Execute selected base learner
Observe Reward
Update Q-Table
?
True
False
Do iteration Until Q-table unchanged or episodes exceed predefined iteration number
Completed Q-table

Proposed Method Results



Yelp review polarity							
Random forest	86.19	82.19	89.70	85.781	90.31	83.16	86.588
Averaging	89.79	93.35	87.28	90.213	86.15	92.77	89.338
Weighted voting	90.64	94.20	88.13	91.064	87.00	93.62	90.189
Majority voting	90.86	94.42	88.35	91.284	87.22	93.84	90.409
(Savargiv et al. 2022)	95.47	99.25	92.98	96.013	91.92	98.21	94.961
(Savargiv et al. 2021)	92.76	96.05	90.78	93.341	88.90	95.19	91.938
Proposed method with initiation	87.93	88.60	87.35	87.971	87.25	88.55	87.895
Proposed method without initiation	98.04	98.71	97.46	98.081	97.36	98.66	98.006
Amazon review polarity							
Random forest	78.25	79.26	77.58	78.411	77.24	78.95	78.086

Code Hands On

https://github.com/itsamirhn/DataMiningProject