Everlasting Iatric Researcher (Eir): Identifying the Article and Reading for Genetic Association Knowledge

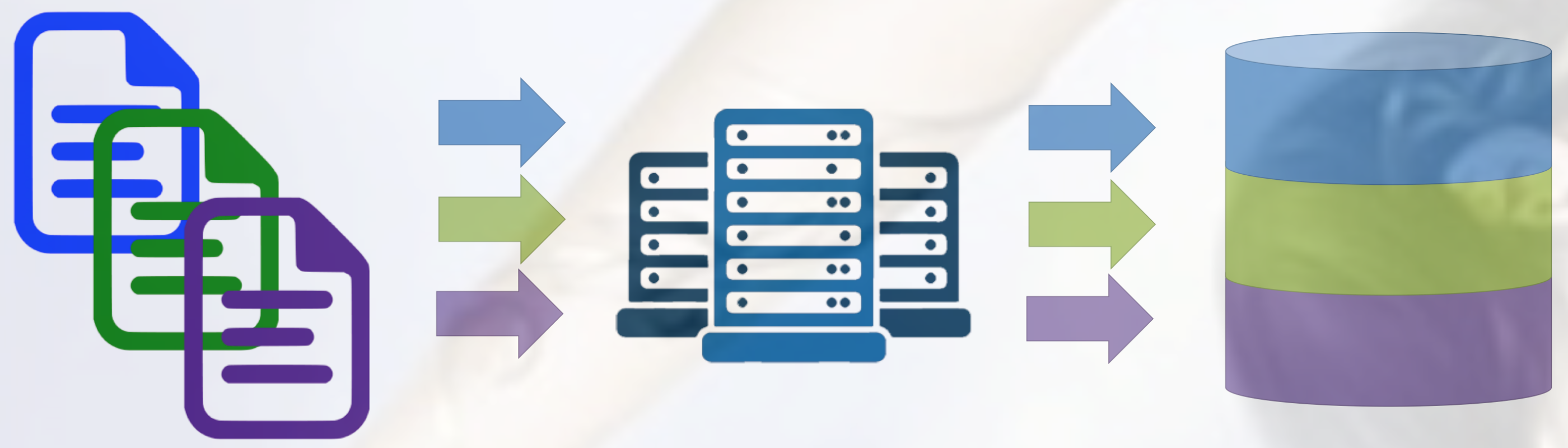
Xiang Liu, Wenting Ye, Haohan Wang

Language Technologies Institute, School of Computer Science, Carnegie Mellon University

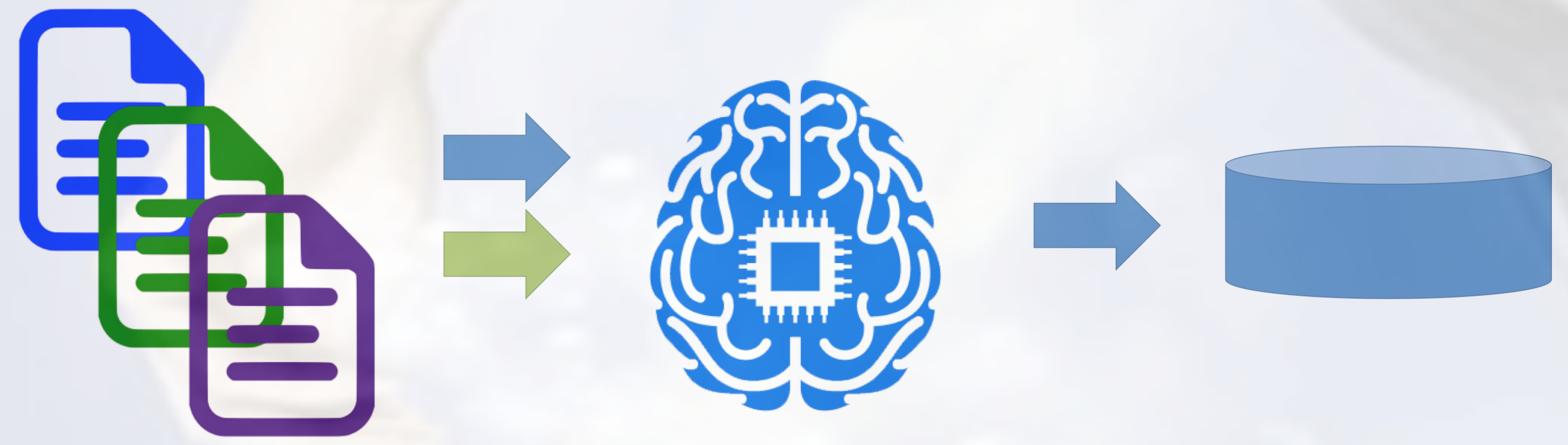
Contribution

- We propose to directly **simulate the behavior of a researcher** (i.e. selecting papers and reading for details) instead of universally examining the entire corpus.
- We develop a deep reinforcement learning model that can select both authentic and informative articles to read.
- We maintain a cutting-edge genetic association relationship database that can be easily queried.

Introduction



- Traditional biomedical text mining learns a massive amount of unreliable knowledge, because the papers themselves are questionable.



- Eir, behaves like human, only reads the paper that she considers trustworthy, and construts knowledge base accordingly.

Deep Reinforcement Learning

Since our model involves a continuous state space *S*, we employ a deep Q-network (DQN) to learn the policy of agent with loss function:

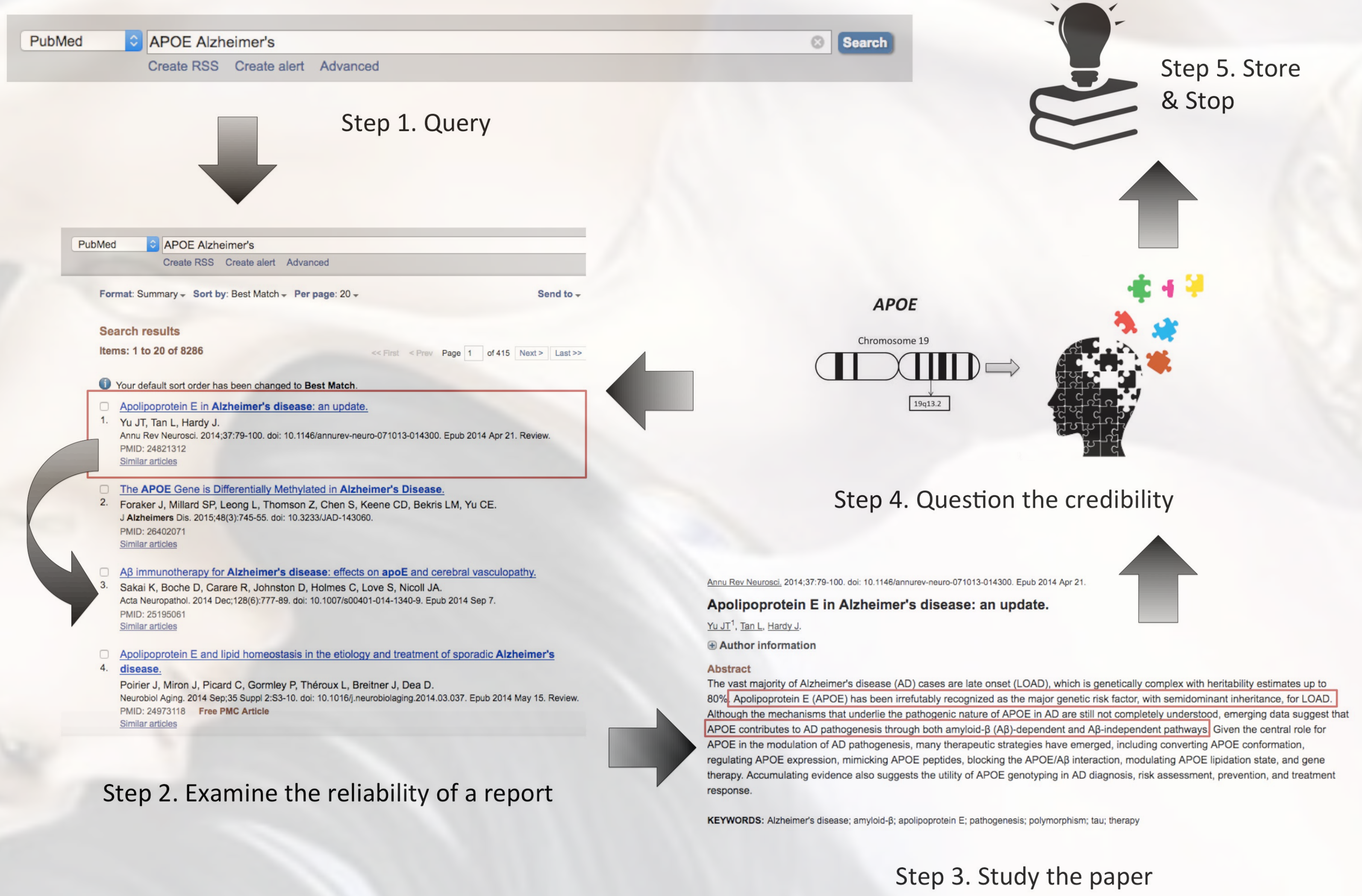
$$L(\theta) = E_{\hat{s},\hat{a}}[(y - Q(\hat{s},\hat{a};\theta))^2]$$

where $y = r + \gamma \max_{a'} Q(\hat{s}', a'; \theta_t)$, and $(\hat{s}, \hat{a}, \hat{s'}, r)$ is selected transition.

- Learn the parameters θ of the DQN using stochastic gradient descent
- Use a (separate) target Q-network to calculate the expected Q-value for "stable update"
- Employ an experience replay memory D to store transitions

Model Framework

Eir's research process is a Markov decision process (MDP), which can be represented as a tuple <*S, A, T, R*>, where *S* = *s* is the space of all possible states, *A* = *a* is the set of all actions, *R (s, a)* is the reward function and *T (s'| s, a)* is the transition function.



PubMed | APOE Alzheimer's | Search
Create RSS Create alert Advanced

Format: Summary Sort by: Best Match Per page: 20 Send to

Search results
Items: 1 to 20 of 8286
<< First < Prev Page 1 of 415 Next > Last >>

Your default sort order has been changed to Best Match.

1. Apolipoprotein E in Alzheimer's disease: an update.
Yu JT, Tan L, Hardy J.
Annu Rev Neurosci. 2014;37:79-100. doi: 10.1146/annurev-neuro-071013-014300. Epub 2014 Apr 21. Review.
PMID: 24821312
Similar articles

2. The APOE Gene is Differentially Methylated in Alzheimer's Disease.
Foraker J, Millard SP, Leong L, Thomson Z, Chen S, Keene CD, Bekris LM, Yu CE.
J Alzheimers Dis. 2015;48(3):745-55. doi: 10.3233/JAD-143060.
PMID: 26402071
Similar articles

3. Aβ immunotherapy for Alzheimer's disease: effects on apoE and cerebral vasculopathy.
Sakai K, Boche D, Carare R, Johnston D, Holmes C, Love S, Nicoll JA.
Acta Neuropathol. 2014 Dec;128(6):777-89. doi: 10.1007/s00401-014-1340-9. Epub 2014 Sep 7.
PMID: 25195061
Similar articles

4. Apolipoprotein E and lipid homeostasis in the etiology and treatment of sporadic Alzheimer's disease.
Poirier J, Miron J, Picard C, Gormley P, Théroux L, Breitner J, Dea D.
Neurobiol Aging. 2014 Sep;35 Suppl 2:S3-10. doi: 10.1016/j.neurobiolaging.2014.03.037. Epub 2014 May 15. Review.
PMID: 24973118 Free PMC Article
Similar articles

Annu Rev Neurosci. 2014;37:79-100. doi: 10.1146/annurev-neuro-071013-014300. Epub 2014 Apr 21.

Apolipoprotein E in Alzheimer's disease: an update.

Yu JT[1], Tan L, Hardy J.

Author information

Abstract

The vast majority of Alzheimer's disease (AD) cases are late onset (LOAD), which is genetically complex with heritability estimates up to 80%. Apolipoprotein E (APOE) has been irrefutably recognized as the major genetic risk factor, with semidominant inheritance, for LOAD. Although the mechanisms that underlie the pathogenic nature of APOE in AD are still not completely understood, emerging data suggest that APOE contributes to AD pathogenesis through both amyloid-β (Aβ)-dependent and Aβ-independent pathways. Given the central role for APOE in the modulation of AD pathogenesis, many therapeutic strategies have emerged, including converting APOE conformation, regulating APOE expression, mimicking APOE peptides, blocking the APOE/Aβ interaction, modulating APOE lipidation state, and gene therapy. Accumulating evidence also suggests the utility of APOE genotyping in AD diagnosis, risk assessment, prevention, and treatment response.

KEYWORDS: Alzheimer's disease; amyloid-β; apolipoprotein E; pathogenesis; polymorphism; tau; therapy

Figure 2: Overview of Eir's possible behaviors

States: The state *s* in the MDP describes the research status of Eir. The information of the state depends on which status Eir is in.

Eir's state	Information
Read the title	• Title • Journal information • Author information
Read the abstract	• Association entity • Confidence • Context information

Table 1: Component of Eir's state in different scenarios

Dataset

- Genetic Association Database (GAD) (Becker et al., 2004)
  - A manually crafted database of 142,000 high quality articles with the association it describes.
- PubMed
  - An online library which contains more than 27 million citations for biomedical literature

Result and Future Work

Observations

At current stage of this project, we obtain the first 4,000 gene-trait associations in GAD with 5,331 articles, and download 35,178 relevant articles from PubMed. The results are showed below.

Entity	Correctness
Gene	83.5%
Trait	76.2%

Figure 3a. Entity correctness

	Title	Abstract
Precision	30.9%	54.6%
Recall	24.6%	46.6%
F_1	0.261	0.50

Figure 3b. Selection accuracy

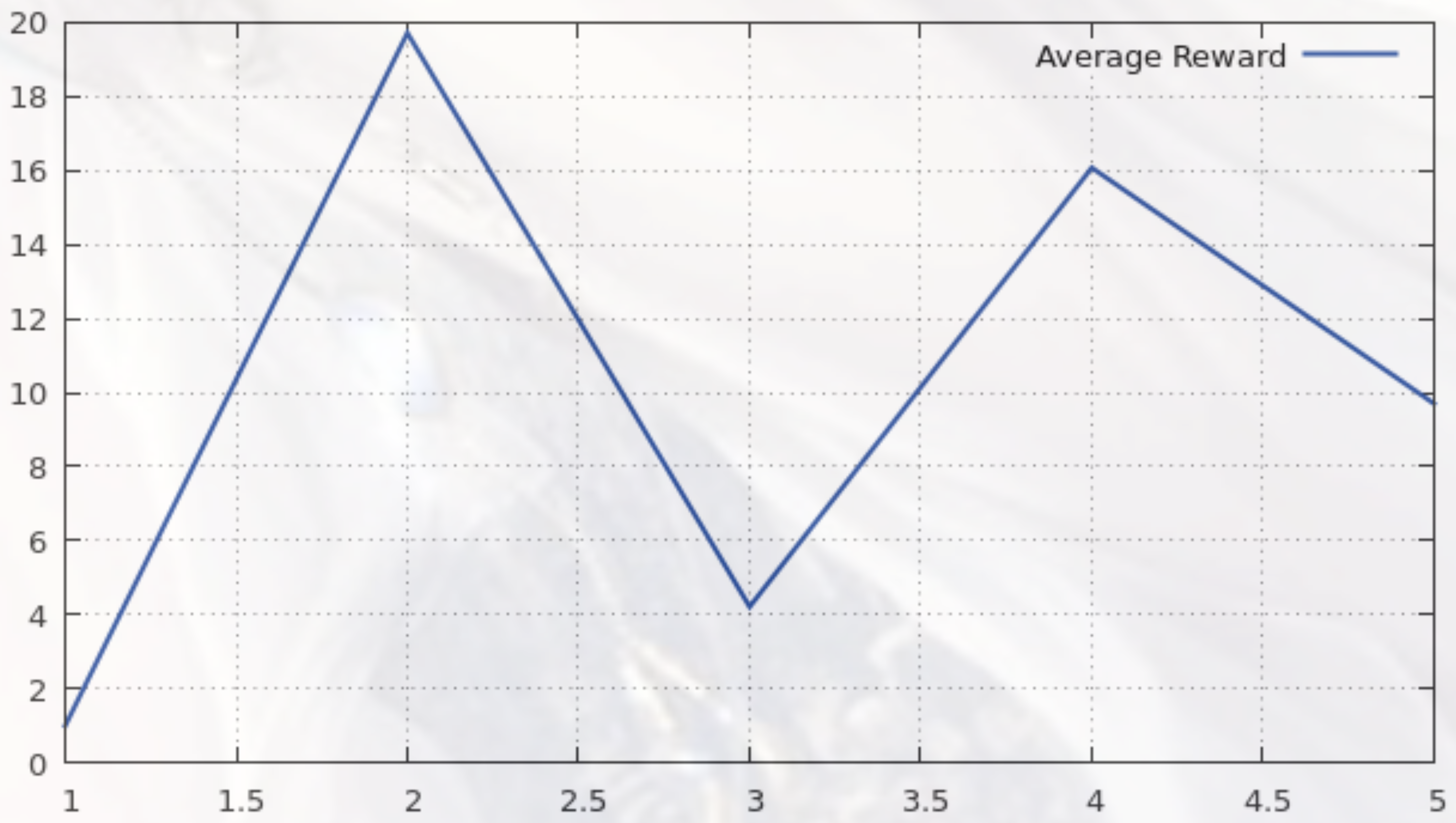


Figure 3c. Average reward

Future works

- A larger amount of text resources
- More powerful text mining tools
- Construction of other biomedical knowledge database, i.e. gene-gene interaction

Contact

- Haohan Wang (haohanw@cs.cmu.edu)

Eir's Origin

Eir is the name of the goddess of medical knowledge and skills in Norse mythology. With current technology, we are hoping to realize the mythology in this modern world.