Selecting Actions for Resource-bounded Information Extraction using Reinforcement Learning

Pallika Kanani
Department of Computer Science
University of Massachusetts, Amherst
Amherst, MA, USA
pallika@cs.umass.edu

Andrew McCallum
Department of Computer Science
University of Massachusetts, Amherst
Amherst, MA, USA
mccallum@cs.umass.edu

ABSTRACT

Given a database with missing or uncertain content, our goal is to correct and fill the database by extracting specific information from a large corpus such as the Web, and to do so under resource limitations. We formulate the information gathering task as a series of choices among alternative, resource-consuming actions and use reinforcement learning to select the best action at each time step. We use temporal difference q-learning method to train the function that selects these actions, and compare it to an online, error-driven algorithm called SampleRank. We present a system that finds information such as email, job title and department affiliation for the faculty at our university, and show that the learning-based approach accomplishes this task efficiently under a limited action budget. Our evaluations show that we can obtain 92.4% of the final F1, by only using 14.3% of all possible actions.

Categories and Subject Descriptors

I.2.6 [**Computing Methodologies**]: Artificial Intelligence—*Learning*

General Terms

Algorithms

Keywords

Resource-bounded Information Extraction, Active Information Acquisition, Reinforcement Learning, Web Mining

1. INTRODUCTION

Resource-bounded Information Extraction (RBIE) is the process of searching for and extracting specific pieces of information from an external information source under a limited budget of resources, such as computation time, network bandwidth, and storage space [5]. The problem of missing information in databases is ubiquitous. However, in many cases, this information is available on some external source, such as the Web. In order to fill in the missing slots, we need a mechanism to automatically extract the required information. In such settings, it is undesirable, infeasible, or even computationally intractable to use traditional information extraction pipelines on the entirety of a vast, external, unstructured or semistructured corpus in order to obtain a relatively small amount of information. Under the RBIE framework, we obtain the required information by issuing appropriate queries to the external source, such as a web search API, downloading only a small fraction of documents from the search results, and processing even fewer of them to extract the specified field.

Consider a real world example. We are building a database of all faculty at a university as shown in Table 1. We have the names, but some of the other information such as contact details, job titles and department affiliations are missing. Surprisingly, in some cases, the university administration does not have such a comprehensive, university-wide database. This may be due to the lack of data exchange, joint appointments across departments, changing contact details, etc. Building such a database would be extremely useful, since it maintains up-to-date records of the faculty, and fosters collaboration across departments. A large portion of this information exists on the Web, but it may not always be found on faculty home pages. Lecturers and faculty in some of the departments do not have home pages, and their information is sometimes scattered around the Web. Finding this information can be challenging, since it is not available in a uniform, structured manner. There are other problems such as name ambiguities and incorrect or incomplete data.

One way to obtain this information is by crawling all the websites under the university domain. This, by itself is a resource-intensive task, since most university websites are large and complex, and we would need to use a lot of computational power to crawl and download the pages, along with the corresponding network bandwidth, and disk space for storing them. We would also lose out on all the information that is scattered on the Web, outside the university domain. Can we accomplish the same task using a much smaller fraction of these resources?

We know that the information missing in the database is available on some relatively small number of pages on the Web. We need to run some extraction algorithms on those pages in order to obtain the required information. But before we can run extraction, we need to download them to our computing infrastructure, and before we can download them, we need to know where they are located on the Web.

WSDM'12, February 8–12, 2012, Seattle, Washington, USA.

Faculty Name	Phone	Email	Job Title	Department Name
Andrew McCallum	(413) 545-1323	mccallum@cs.umass.edu	Professor	Computer Science
Jerrold S. Levinsky	?	?	Lecturer	Legal Studies
Edward G. Voigtman	?	?	?	?
Robert W. Paynter	?	?	?	Anthropology

Table 1: Example Database of University Faculty

A search engine API, such as Google can help us retrieve these web pages. We first formulate queries driven by information that is already available in the database, issue them to the search interface, obtain the location of the web pages, and download them. Then we can run the necessary algorithms to extract information, and use it to fill missing entries in the database. This process is more efficient than indiscriminate processing, and would use relatively smaller amounts of resources.

The problem of Resource-bounded Information Extraction (RBIE) was first introduced in our previous work [5]. Queries are formed by combining existing, relevant information in the database with user defined keywords. All such queries are issued to the search API, and all of the result documents are downloaded. The resource savings come from selecting a subset of the web documents to process by exploiting the network structure in the data. In general, we may need multiple queries to obtain information about a single entry in the database, and some queries work better than others. In our university faculty example, we may form different queries with keywords such as "curriculum vitae" or "home page", and it may be the case that one of them is often more successful than the other in finding the information we need. In some cases, the information in different fields may be interdependent, and finding one before another may be more efficient. In order to make the best use of available resources, we need to issue the most effective queries first.

In most scenarios, one only need process a subset of the documents returned by the queries. We need to know which of the search results are most likely to contain the information we are looking for. Information returned in the search result snippet can be exploited to decide if a web page is worth downloading. Similarly, some preliminary observation of the downloaded document can be useful to decide if it is worth passing through an expensive extraction pipeline. Hence, instead of viewing RBIE as selecting a subset of documents to process, we cast the information-gathering task as a series of resource-consuming actions, along with a mechanism to select the best action to perform at each time step.

In this paper, we formulate the RBIE problem formally as a Markov Decision Process (MDP), and propose the use of reinforcement learning techniques for solving it. The *state* of this MDP is the state of the database at each time step, and *action* is any act that leads to obtaining the required information, such that performing the action in one state leads to a different state. RBIE process is then finding the optimal policy in this MDP, so as to obtain most information with the given budget of actions, since we assume that actions consume resources. In RBIE from the Web context, actions are *query*, which is issuing a query to a search API, *download*, which is downloading a web document, and *extract*, which runs an actual extraction algorithm on a document. We assume uniform cost for each type of action in this paper, but the proposed framework can easily be extended by incorporating a specific cost model for the actions and assigning the budget accordingly. By formulating RBIE for the Web as an MDP, we can explore the rich methods of optimal action selection offered by reinforcement learning.

In the RBIE for the web setting, query actions might not lead to immediate reward, but they are necessary to perform before download and extract actions, which may lead to positive rewards. Hence, we need a method that models delayed rewards effectively. We propose the use of temporal difference q-learning to learn the value function for selecting the best action from a set of alternative actions, given a certain state of the database. In our preliminary work [4], we explored a fast, online, error driven algorithm, called SampleRank to learn this value function. Since both SampleRank and q-learning are novel approaches for the RBIE framework, we compare their relative performance on the task of finding emails, job titles and department affiliation for faculty in our university.

In general, we can use any model of choice for information extraction in our framework that can extract the required information from a web page, and provide a confidence score for the extracted value. This score can be used to choose the best among the potential candidate values, and to determine whether or not an existing entry in the database should be updated by the newly extracted value. We present a simple, but novel information extraction method that can easily scale to large problem domains. The basic idea is to generate a list of potential candidate values from the web page, and using a binary classifier, such as maximum entropy, to classify them as being correct values or not, by observing features of the context in which they are found. The candidate with the maximum probability of being correct is used to fill the entry in the database.

Our experiments show that the q-learning strategy performs better than the baseline action selection strategies, as well as SampleRank based approach to learning a value function. Given the large number of actions to choose from at each time step, and the size of the corresponding state space, the policy learned is impressive. The q-learning based approach is able to obtain 92.4% of the final F1, by only using 14.3% of all possible actions, demonstrating the effectiveness of our method.

2. RESOURCE-BOUNDED INFORMATION EXTRACTION

2.1 General Problem Definition

The general problem of Resource-bounded Information Extraction (RBIE) is defined as follows. We are given a database with missing values in some of the entries, and a set of possible actions to help acquire that information from an external source, such as the Web. Select the best action from the set of alternative actions available at each

time point, so as to acquire most information with the given budget on the number of actions.

2.2 Resource-bounded Information Extraction From the Web

For RBIE from the Web, we consider three different types of actions - *query*, *download*, and *extract*. A *query* action consists of issuing a single query to a web search API and obtaining a set of search results. In order to form the query, we need to use some existing information from an input record in the database and a set of keywords. A *download* action consists of downloading the web page corresponding to a single search result. Finally, an *extract* action consists of performing extraction on the downloaded webpage to obtain the required piece of information and using it to fill the slot in the original database. Note that each instantiated 'action' consists of the type of action as well as its corresponding argument, namely, what query to send for which instance, which URL to download, or what page to extract.

In the case of RBIE from the Web, the *query* actions can be initialized at the beginning of the task because we know which instances have missing fields, and the types of queries that can be used; but *download* actions and *extract* actions are generated dynamically and added to the list of available actions. That is, after a *query* action is performed, the *download* action corresponding to each of the search results is generated. Similarly, after a web page is downloaded, the corresponding *extract* action is generated. At each time point, only the actions that are instantiated can be considered as alternative valid actions to be performed. The RBIE task is to select the "best" action at each time point from a set of all valid actions.

Before selecting an action to perform at each time step, we need to consider several factors. We need to take into account the current state of the database, such as the number of slots filled and the uncertainty about them. We need to take into account the context provided by the intermediate results of all the actions so far, such as the results of the queries, pages that are not yet downloaded and processed. Even if this context is not yet in the database, these intermediate results can provide valuable information for deciding which action to select. Finally, we also need to consider the properties of the candidate action itself, before selecting it.

We assume that we are given an existing model, M_e for extracting the required pieces of information from a single web page. We also assume that this model provides a confidence score for each value predicted. This score can be used to choose the best among the potential candidate values, and to determine whether or not an existing entry in the database should be updated by the newly extracted value.

2.3 Markov Decision Process Formulation

In this paper, we cast the Resource-bounded Information Extraction problem as solving a Markov Decision Process (MDP), M, where the states represent the state of the database at a given time, along with any intermediate results obtained from the Web, and actions represent the *query*, *download*, and *extract* actions as described in the previous section. We represent state as a tuple $S_t\langle DB_t, I_t, I'_t\rangle$, where DB_t is the state of the database at time t, I_t is the list of intermediate URL results and I'_t is the list of intermediate page results obtained till time t. The MDP for RBIE is described as a tuple, $M\langle S_0, \gamma, T(S, a, S'), R(S)\rangle$, where S_0 is the initial state of the database, γ is the discount factor, $T(S, a, S')$ is the state transition probability, or the probability that action a in state S at time t will lead to state S' at time $t+1$, and $R(S)$ is the reward function for being in state S.

Algorithm 1 Resource-bounded Information Extraction for the Web Using Q-function

Input:
Database DB with missing entries, E_i
Learned Q-function $Q(a, S)$
Learned extraction model, M_e
Time budget, b
Initialize all queries using keywords
$t = 0$
while $t <= b$ **do**
  $a_{t+1} = \arg\max_a Q(a, S)$
  if a_{t+1} is a *query action* **then**
    Issue query to a web search API
    Enqueue corresponding *download actions*
  else if a_{t+1} is a *download action* **then**
    Download the web page
    Enqueue corresponding *extract action*
  else if a_{t+1} is an *extract action* **then**
    Extract all candidate values from the web page
    Score each candidate using the model, M_e
    Fill the value of the best candidate in E_i
  end if
  $t = t + 1$
end while

One of the standard ways of solving an MDP is q-learning[15], which provides a way to learn to select the best action at each time step. The Q-function, $Q(a, S)$ is the expected utility of taking action a in state, S. Hence, the best action to select at each step is:

$$a_{t+1} = \arg\max_a Q(a, S) \tag{1}$$

Algorithm 1 summarizes the RBIE for Web framework for filling missing information in a database using Q-function.

2.4 Learning Q-function

At the heart of solving an MDP for RBIE is the Q-function, $Q(a, S)$. We now discuss a method to learn it from real data. Much of the material in this section follows from [12, 14].

We know that Q-function obeys the following constraints:

$$Q(a, S) = R(S) + \gamma \sum_{S'} T(S, a, S') \max_{a'} Q(a', S') \tag{2}$$

To use this update equation, we need to learn the transition probability model, $T(S, a, S')$, which is difficult in our setup. Hence, we use the temporal-difference, or TD q-learning approach, which is also called model-free, because it lets us learn the Q-function without using the transition probability model. The update equation for TD q-learning is [1]:

$$Q(a, S) \leftarrow Q(a, S) + \alpha(R(S') + \gamma \max_{a'} Q(a', S') - Q(a, S)) \tag{3}$$

[1]There is some disagreement amongst q-learning experts about which form of reward function to use. We choose to use $R(S')$.

Algorithm 2 Temporal difference q-learning for RBIE, with ϵ-greedy exploration

Input:
Training database, DB
Initial parameters, θ
Q Function, $Q_\theta(a, S) = \sum_i \theta_i \phi_i(a, S)$
Reward Function, $R(S)$
Learning Rate, α
Discount factor, γ
$S_0 \leftarrow$ Initial State of DB
for $t \leftarrow 0$ to number of iterations T **do**
  $\epsilon = 1 - \frac{1}{T}$
  With probability ϵ, pick a random action, a_t
  With probability $1 - \epsilon$, pick $a_t = \arg\max_a Q_{\theta_t}(a, S_t)$
  $S_{t+1} = a_t(S_t)$ //perform a_t
  Let a' be all the valid actions from state, S_{t+1}
  for $i = 0$ to number of features **do**
    $\theta^i_{t+1} = \theta^i_t + \alpha[R(S_{t+1}) + \gamma \max_{a'} Q_{\theta_t}(a', S_{t+1}) - Q_{\theta_t}(a_t, S_t)]\phi_i(a_t, S_t)$
  end for
end for

Where, α is the learning rate. For any real-world RBIE task, the state space for the corresponding MDP is large enough to make it very difficult to learn this function accurately. Hence, we use function approximation. We represent the Q-function as a weighted combination of a set of features as follows:

$$Q_\theta(a, S) = \sum_i \theta_i \phi_i(a, S) \tag{4}$$

Where $\phi_i(a, S)$ are the features of the state S and action a, and θ_i are the weights on those features that we wish to learn. We now use the following equation [12, 14] for updating the values of θ_i to try to reduce the temporal difference between successive states.

$$\theta_i \leftarrow \theta_i + \alpha \left[R(S') + \gamma \max_{a'} \hat{Q}_\theta(a', S') - \hat{Q}_\theta(a, S) \right] \frac{\partial \hat{Q}_\theta(a, S)}{\partial \theta_i} \tag{5}$$

We can now use this update equation to learn the parameters of our Q-function from training data. The TD-q-learning algorithm for RBIE is described in Algorithm 2. Note that we use ϵ-greedy approach for exploring the state space, where ϵ decreases in proportion to the number of training iterations.

We also need to design a custom reward function, $R(S)$ for using this algorithm. Under the RBIE from the Web setting, we can compute value of the reward function after performing action a_{t+1} on $S_t = \langle DB_t, I_t, I'_t \rangle$ as a weighted sum of correct, incorrect and total number of filled values, number of correct candidates found, and some properties of the intermediate results.

2.5 Building the Extraction Model

In general, we can use any model of choice for information extraction in our framework that can extract the required information from a web page, and provide a confidence score for the extracted value. In this section, we present a simple, but novel information extraction method that can easily scale to large problem domains. The basic idea is to generate a list of potential candidate values from the web page, and using a binary classifier, such as Maximum Entropy, to classify them as being correct values or not, by observing features of the context in which they are found. Algorithm 3 describes how we train the model.

Let E be the set of entries with missing values in the database. We use patterns and lexicons to generate a list of candidates, C_{E_i} for each entry, $E_i \in E$. A *candidate* is a unique string that is a potentially correct value for an entry in the database. Each candidate, $c_j \in C_{E_i}$, consists of a list of *mentions*, M_j, which represent the actual occurrence of the candidate string in the web documents. Each candidate may have multiple mentions, across different web pages. Corresponding to each mention, $m_k \in M_j$, we have a list of properties, or features, $f(m_k)$ which describe the context in which it was found. Since we are interested in classifying the single, canonical value of these mentions, i.e, the candidate, we collapse the properties of different mentions for a candidate c_j into a single feature function, $f(c_j)$.

Let y_{ij} be a binary variable that represents whether c_j is the correct value for entry E_i. We can then represent the probability of c_j being the correct candidate as:

$$P(y_{ij}|c_j) = \frac{1}{Z} \exp(\sum_l \lambda_l f_l(c_j, y_{ij})) \tag{6}$$

Where, λ_l are the weights on the features, and Z, the normalization factor is given by:

$$Z = \sum_y \exp(\sum_l \lambda_l f_l(c_j, y_{ij})) \tag{7}$$

Since this is a supervised approach, our training data consists of the true values of E, which can be used to train the classifier. At test time, during an extract action, we classify each $c_j \in C_{E_i}$ at that time point, and select the one with the maximum posterior probability, $P(y_{ij}|c_j)$, as the "best" candidate to fill the slot.

3. SAMPLERANK FOR RBIE

In our preliminary work, we explored a different approach to learning a value function for selecting actions for a different problem domain[4]. We use a fast, online, error driven learning algorithm, called SampleRank [1, 16]. Since, we are also interested in investigating the effectiveness of the SampleRank approach in our current problem domain, we give its brief introduction here. For further details on training SampleRank for RBIE, refer to [4].

Remember that we represent a state as, $S_t\langle DB_t, I_t, I'_t \rangle$. Our goal is to learn a value function similar to the Q-function, called $V(DB_t, I_t, I'_t, a)$, which is used to select the best action in a given state. In order to learn this function from training data, we first assume that its functional form is as follows:

$$V(DB_t, I_t, I'_t, a) = \exp(\sum_k \lambda_k \phi_k(DB_t, I_t, I'_t, a)) \tag{8}$$

Where, λ_k are model parameters and ϕ_k are feature functions, defined over the database context, the current action, and the results of all previous actions.

We start training with state S_0, that represents the original state of the database. We consider all available actions at this point, and sample from states that result from these actions. We choose the state S^*, which is the result of the best action $a*$, predicted by V, and the state S', which is the best state predicted by the objective, or reward function, $R(S)$. SampleRank is an error driven learning algorithm, which

Algorithm 3 Building Extraction Model, M_e

Input:
Training Database DB with entries, E
Pattern or Lexicon Matcher, $L(w)$ that returns a set of matches, M_w from a Web Page, w
Feature functions, $f(.)$ describing context of M_w
A Supervised Learning algorithm, such as Max Ent
Initialize all queries using keywords
Initialize set of potential candidates per entry, $C_{E_i} = \{\}$
Initialize set of candidates for training, $C_t = \{\}$
while Any more actions remain **do**
  Pick a random action, a to perform
  if a is a *query action*, or a *download action* **then**
    Perform a and enqueue corresponding *download* or *extract actions*
  else if a is an *extract action* for Web Page, w **then**
    $M_w \leftarrow L(w)$
    for Each match, $m_k \in M_w$ **do**
      if String value (m_k) matches $c_j \in C_{E_i}$ **then**
        Add m_k as a mention of c_j
        Merge the features, $f(m_k)$ with $f(c_j)$
      else
        Create a new candidate, c_j, and add to C_{E_i}
        $label(c_j) \leftarrow$ string value (c_j) = true value (E_i)?
        Add m_k as a mention of c_j
        Set $f(c_j) \leftarrow f(m_k)$
      end if
    end for
  end if
end while
for all C_{E_i} **do**
  $C_t \leftarrow C_t \cup C_{E_i}$
end for
$M_e \leftarrow$ Train a Max Ent classifier with $f(c_t)$, for $c_t \in C_t$

lets us update parameters when the function learned up to a given point makes a mistake. We say the ranking is *in error* if the function learned so far assigns a higher score to the sample with the lower objective, or reward value, $R(S)$, i.e.:

$$[(V_\Lambda(S^*) > V_\Lambda(S')) \wedge (R(S^*) < R(S'))] \vee [(V_\Lambda(S^*) < V_\Lambda(S')) \wedge (R(S^*) > R(S'))] \quad (9)$$

When this condition is true, we update the parameters, Λ using perceptron update as follows:

$$\Lambda^t \Leftarrow \Lambda^{t-1} + \alpha(\phi(S'_t, a'_t) - \phi(S^*_t, a^*_t)) \quad (10)$$

where α is the learning rate used to temper the parameter updates. Note that SampleRank is a special case of reinforcement learning, which makes the comparison between the two approaches very interesting.

4. RELATED WORK

4.1 Resource-bounded Reasoning

Knoblock et al. [7] introduced the idea of using planning for information gathering, followed by the development of resource-bounded reasoning techniques by Zilberstein et al. [18]. Elliasi-Rad [2] explored the problem of building an information extraction agent, but did not address the problem of acquiring specific missing pieces of information on demand. The problem of Resource-bounded Information Extraction (RBIE) was first introduced in our previous work [5], in which the main idea was to select a subset of the web documents to process by exploiting the network structure in the data. The example task in [5] is to find a missing year of publication in citation data. All available queries are issued, and all the search results are downloaded, which are then filtered using a simple heuristic. The number of documents that need to be processed for extraction is reduced by propagating information obtained from the Web through the underlying citation graph structure. In our preliminary work[4], we explored the task of finding URLs of the faculty directory pages of top Computer Science departments in the U.S., and used SampleRank for learning the value function.

4.2 Information Extraction From the Web

In the traditional information extraction settings, we are usually given a database schema, and a set of unstructured or semi-structured documents. The goal of the system is to automatically extract records from these documents, and fill in the values in the given database. These databases are then used for search, decision support and data mining. In recent years, there has been much work in developing sophisticated methods for performing information extraction over a closed collection of documents. Several different approaches have been proposed for different phases of information extraction task, such as segmentation, classification, association and coreference. Most of these proposed approaches make extensive use of statistical machine learning algorithms, which have improved significantly over the years. However, only some of these methods remain computationally tractable as the size of the document corpus grows. In fact, very few systems are designed to scale over a corpus as large as, say, the Web [3].

Among the large scale systems that extract information from the Web are KnowItAll [3], InfoSleuth [11] and Kylin [17]. The goal of the KnowItAll system is a related, but different task called, "Open Information Extraction." In Open IE, the relations of interest are not known in advance, and the emphasis is on discovering new relations and new records through extensive web access. In contrast, in our task, we are looking for specific information and the corresponding schema is known. The emphasis is mostly on filling the missing fields in known records, using resource-bounded web querying. InfoSleuth focuses on gathering information from given sources, and Kylin focuses only on Wikipedia articles.

The Information Retrieval community is rich with work in document relevance (TREC). However, traditional information retrieval solutions can not directly be used, since we first need to automate the query formulation for our task. Also, most search engine APIs return full documents or text snippets, rather than specific values. A closely related family of methods is question answering systems. These systems do retrieve a subset of relevant documents from the Web, along with extracting a specific piece of information. However, they target a single piece of information requested by the user, whereas we target multiple, interdependent fields of a relational database. QA usually interprets natural language question, whereas we need a keywords based mechanism for formulating queries. Most QA systems do not focus on prioritizing information acquisition actions, and the ideas in this paper could prove useful in building them. The seman-

tic web community has been working on similar problems, but their focus is not targeted information extraction.

4.3 Active Information Acquisition

Learning and acquiring information under resource constraints has been studied in various forms. *Active learning* selects the best instances to label from a set of unlabeled instances; *active feature acquisition* [10] explores the problem of learning models from incomplete instances by acquiring additional features; *budgeted learning* [8] identifies the best set of acquisitions, given a fixed cost for acquisitions. At test time, the two common scenarios are selecting a subset of features to acquire, e.g. [13], and selecting the subset of instances for which to acquire features [6]. In our work, the resource-constraints are only applied at test time, and availability of unlimited resources is assumed at training time.

5. EXPERIMENTS

5.1 Extracting Faculty Information

Given a list of names of university faculty, our goal is to extract their email address, job title and department affiliation from the Web. In this section, we describe how we apply the RBIE framework to build a system that can efficiently acquire this information. We also describe experiments testing the effectiveness of SampleRank and q-learning algorithms in selecting the most effective actions at each time step.

This is a challenging task due to several factors. In some cases, this information is readily available on faculty home pages, which are semi-structured. However, lecturers and faculty in many departments do not have home pages. Their information is scattered around the Web, without a uniform structure. Web pages are noisy, and may lead to unexpected errors while performing extraction. Name ambiguity is another challenge, since many of the faculty have common names they share with other famous personalities. Some information on the Web is stale, or contradicting. For example, a faculty member can be listed on one page as "assistant professor", while on another as "associate professor", reflecting a recent change of title. Finally, some information is not available on the Web at all.

5.2 Dataset Description

We start with a list of faculty from University of Massachusetts at Amherst. We randomly choose 100 of these records as our dataset. The fields contain the first, middle and last name of the faculty, their email address, a list of job titles, and a list of department affiliations. Joint appointments lead to multiple job titles and department affiliations. The dataset we received contains several inaccuracies and is cleaned for better evaluation of our methods. For example, in some cases, a single column contains names of different departments. These are split into multiple columns. Punctuation and abbreviations, such as "Assoc. Prof." are cleaned and expanded. Despite the cleaning effort, the dataset we use is incomplete and contains errors. For example, the most current job titles are not reflected, and only one email address is included in the dataset, which may not be the one used by the person, or published on the Web. These imperfections in the data make both training and evaluation of our system challenging. Another problem in evaluating the accuracy of our system is the "generic-specific" problem in department names. For e.g., our system might predict the department affiliation for a faculty as "finance", while it might be listed as "management" in the ground truth dataset, or vice-versa. Since we use exact string match, we may even miss a match such as "school of management". Despite the difficulties, it is an interesting real world task for RBIE.

Dataset	# Faculty	# Queries	# Docs	Total Actions
Training	70	1400	13686	28772
Testing	30	600	6065	12730
Total	100	2000	19751	41502

Table 2: Datasets

Name	Name + Univ
Name in quotes	Name in quotes + Univ
Name In Univ	Name in quotes In Univ
Name w/ middle In Univ	Name w/ middle + Univ
Name + CV	Name + Univ + CV
Name + "Resume"	Name + Univ + "Resume"
Name + "Profile"	Name + Univ + "Profile"
Name + "Bio"	Name + Univ + "Bio"
Name + HomePage	Name + HomePage In Univ
Name + "Contact"	Name + "Contact" In Univ

Table 3: Types of Queries. 'Name' : first and last name, 'CV' : "curriculum vitae", 'Univ' : "university of massachusetts at amherst" and 'In Univ' : "site:umass.edu"

We use the Google search API for our experiments. In our task, the three fields that we extract are related to each other and often found in the proximity of each other on the same web pages. Hence, our query actions correspond to the entire record in the database, as opposed to a single 'entry', or cell. We formulate 20 different types of queries per faculty, as shown in Table 5.2, and consider top 20 hits returned by the search API. Assuming that we are not operating under resource-constraints, i.e., we perform all possible actions available, we get the dataset as described by Table 2. This table also helps estimate the size of the state space for our problem.

5.3 Training the Extraction Models

Before we move to the action selection experiments, we must build a model for extracting the relevant fields from individual web pages. Section 2.5 describes the algorithm we use for training the model. We use the MALLET [9] toolkit's implementation of the maximum entropy classifier. The available data is first split by 70%-30% for training and testing. The training phase for the extraction model is not resource-constrained, i.e., we use all possible query, download and extract actions.

The algorithm described in section 2.5 uses a pattern or lexicon matcher that returns a set of matches from a web page. These matches are added as a list of candidates to be filled in the database entry. For emails, we use a regular expression to match all the emails found in the web document. For job titles and department affiliations, we first build N-grams from body of the web page, where $N = 1, 2, 3, 4$. These N-grams are matched against lexicons to find candidate mentions in the web page. The features used to describe the context of these matches are shown in Table 5.3. The features across a mention are collapsed by using an 'OR' operator, since they are mostly binary. That is, if any feature is turned on in one of the mentions, it would be turned on

Features for Email Extractor
Type of query used
Email domain is from UMass
Web page domain name from UMass
Email host and web page URL host match
Relative position of faculty name and email
Match between faculty name and email username
Similarity between faculty sname and email username
Features for Job Title Extractor
Too many matches found on page
Web page domain name from UMass
Web page URL contains faculty name
Position of match on the document
The words "Assistant" or "Associate" preceds match
Relative position of faculty name and job title
Features for Department Extractor
Too many matches found on page
Web page domain name from UMass
Web page URL contains faculty name
Position of match on the document
The word "Department" precedes match
Relative position of faculty name and department name

Table 4: Features of the Extraction Models

Measure	Email	JobTitle	Department
Accuracy	97.97	92.50	96.94
Yes Precision	77.78	36.36	54.54
Yes Recall	87.50	15.38	44.44
Yes F1	82.23	21.62	48.97
No Precision	99.28	94.15	98.11
No Recall	98.57	98.06	98.73
No F1	98.93	96.06	98.42

Table 5: Performance of the Extraction Models

for the candidate. In our early experiments, we found this method perform better than other merging operations, however, in future, we can build a more sophisticated method.

Let us first study the performance of the candidate classifier, in isolation of the resource-bounded information extraction task. Any inaccuracy in this model, will not only result in poor accuracy during the RBIE process, but also misguide it due to inaccurate confidence prediction. Table 5 shows the classification performance of M_e. Note that F1 is the harmonic mean of precision and recall. The main reasons for relatively lower F1 values on this model are inaccuracy of training data as described above, as well as the noisy nature of the web data. The advantage of using this model is that it is easy to build, and is scaleable for very large scale problems. In the future, we would like to experiment with a more sophisticated extraction model, in order to facilitate better accuracy of the classifier, as well as the RBIE process.

5.4 Evaluation

At test time, we start with the database that contains names of faculty. All other columns are empty. We consider this as time, $t = 0$. We assume that each action takes one time unit. The action selection scheme that we are testing selects one of the available actions, which is performed as described in Algorithm 1. The action is then marked as completed and removed from all available actions. If an extraction action is selected, it may affect the database by filling a slot and altering the confidence value associated with that slot. We evaluate the results on the database at the end of a given budget, b, or if we run out of actions.

We are interested in finding the email address, job title and department affiliation, all of which can have multiple true values. Note that this also includes minor variations. Throughout our evaluations, we compare against the multiple values of a column, and declare a match if the predicted value matches at least one of them. We use the following evaluation metrics to measure our system's performance. Since our task is slightly different from a traditional information extraction task, we use the following definitions of evaluation metrics. Note that extraction recall measures the proportion of entries for which a true candidate value has been extracted from the web page. It may or may not get ranked as the "best" candidate. However, for the purpose of evaluating the order of selecting the query, download and extract actions, this is a very important metric. Even though at test time, it is independent of the performance of the underlying extraction model, M_e, it is still influenced indirectly by M_e through training.

$$\text{Precision} = \frac{\text{No. of Correctly Filled Entries in the Database}}{\text{No. of Filled Entries in the Database}}$$

$$\text{Recall} = \frac{\text{No. of Correctly Filled Entries in the Database}}{\text{No. of Test Entries in the Database}}$$

$$\text{Extraction Recall} = \frac{\text{No. of Correct Candidates Extracted}}{\text{No. of Test Entries in the Database}}$$

$$\text{F1} = \frac{2*\text{Precision}*\text{Recall}}{(\text{Precision}+\text{Recall})}$$

5.5 Baselines

We use two baselines for our experiments : random, and straw-man. At each time step, the random approach selects an action randomly from all available actions. The straw-man approach is actually an extremely strong competitor and works as follows. The first query in the list is issued for each test instance. Next, the first hit from the search result for each test instance is downloaded and processed for extraction. Then, subsequent hits from the search result for each test instance are downloaded and processed. Finally, subsequent queries are issued in the descending order, followed by their corresponding download and extract actions. Note that this approach quickly fills up the slots with the top hits of the queries, making it a very difficult baseline to beat.

5.6 Learning Q-function from Data

We now describe how parameters θ for Q-function $Q_\theta(a, S)$ are learned using training data. Table 5.6 shows the features used. Note that at train time, we do not impose resource constraints. That is, training is performed till more actions are available. However, we only run Q-function parameter updates for a given number of iterations, which acts as a type of budget. We determine the number of iterations and learning rate empirically.

Similar to the test time, we start with a database with the email, job title and department name columns empty. The true values of these columns are used only to calculate the reward function. We initialize the parameters to zero. At each time step, we explore all possible actions, and update the parameters as described in Algorithm 2. We then choose the next action to perform as per the updated parameters and proceed similarly for the specified number of iterations.

Features related to query, download and extract actions
Type of query
Features related to download and extract actions
Hit value in the search result URL is from UMass Webpage is HTML Title contains keywords Title contains faculty name
Features related to extract actions
Appropriate Size Bad request code found

Table 6: Features for learning using SampleRank and Q-function

Frac. of Action Budget	Frac. of Best Extraction Recall
00.00%	00.00%
14.29%	77.78%
28.57%	93.65%
42.86%	96.83%
57.14%	96.83%
71.43%	100.00%
85.71%	100.00%
100.00%	100.00%

Table 7: Effectiveness of Biased-Q-learning in obtaining extraction recall over total entries using an oracle

We use the following reward function for training. Here, n is the number of slots filled in the database, d is the number of slots filled correctly, $\bar{d}$ is the number of slots filled incorrectly, m is the number of correct candidates found, and C_n, C_d, $C_{\bar{d}}$, and C_m are the corresponding coefficients. In our initial experiments, using intermediate results was not helpful.

$$R(S_{t+1}) = C_n * n + C_d * d + C_{\bar{d}} * \bar{d} + C_m * m \qquad (11)$$

Since we are interested in comparing the SampleRank approach, we use the same features and reward function as q-learning. We also introduce a variation of q-learning, where the policy is initialized using the straw-man approach, followed by the normal q-learning. In this case, a bias value proportional to the rank of actions proposed by the straw-man method is added to the q-function value for each state-action pair. We call this method 'biased-q-learning' in our experiments.

5.7 Results and Discussion

We now compare the test-time performance of the two baselines, the value function learned using SampleRank, and the Q-function on their ability to select good actions at each time step. Note that we have already evaluated performance of the extraction method, and we are now focussing on quality of the action selection strategy. We evaluate performance after each 2000 actions from 0 to 14000 (since the total number of actions at test time is 12730). The most effective action selection scheme is the one that is fastest in achieving high values of evaluation metrics.

5.7.1 RBIE Using an Oracle

We first evaluate performance of the four action selection schemes in the presence of an oracle that perfectly classifies each candidate as the correct value for an entry or not with infinite confidence. We do this to isolate the effect of inaccuracies in the extraction model, M_e, which can severely misguide the RBIE system with wrong confidence values. For e.g., even if the action selection scheme selects a good web page for extraction, M_e can choose the wrong candidate for updating the value in the corresponding slot. While training the Q-function, this translates to incorrect reward values, which can severely impede learning. Table 5 shows that the F1 value for 'yes' label for each of the extractors are not high enough to avoid these problems.

Figure 1 shows the extraction recall values during the RBIE process for different fields, and the total number of entries. Note that in the presence of an Oracle, other evaluation metrics are not useful, since the precision is always 1, and the recall is the same as the extraction recall. We ran 2000 training iterations of SampleRank, q-learning, and biased-q-learning, with learning rate, $\alpha = 0.005$, $C_n = 0.01$, $C_m = 1$, $C_d = 0.1$, $C_{\bar{d}} = -0.05$, and a discount factor, $\gamma = 0.9$ for this experiment.

As we can see, the straw-man method is extremely effective, because it knows to process the top hits for a good query for each entry first. Given the complexity of the action domain, and the size of the state space, this policy is very difficult for a Q-learner to learn. It does, however, learn to beat the random action selection, as well as the value function learned by SampleRank. As expected, both q-learning and biased-q-learning perform better than SampleRank due to their modeling of delayed rewards, despite the use of exactly the same features and reward functions. In the example of extracting emails, biased-q-learning outperforms all other methods, while q-learning outperforms other methods over total entries.

To gain some intuition about the policy learned by q-learning, let us look at a few top features for each action type. For *query* actions : query type with just the name of the faculty, query type with the name of the faculty and the keyword, 'Homepage'. For *download* and *extract* actions : the URL is html, URL is from a umass.edu domain, the title contains faculty name, and combinations of the corresponding query type and range of hit values. For *extract* actions, it also learned high weights for features that looked at the size of the documents. Hence, even though the straw-man method has the advantage of background human knowledge, such as the importance of hit value in the search engine results, the learning-based methods learn new and more elaborate patterns that show relative usefulness of different types of queries and help identify promising documents to process by 'examining' them.

Table 7 presents the percentage of the best extraction recall obtained over total entries for the fraction of the action budget. For example, 77.78% of the best achievable extraction recall (using all available actions) is obtained using only 14.29% and so on. This shows that the proposed RBIE methods can be effective in obtaining most of the useful information using only a fraction of the resources. One thing to note here is that even with using oracle we are not able to achieve perfect extraction recall at the end of al available actions. This illustrates how challenging the problem of finding information about people online is.

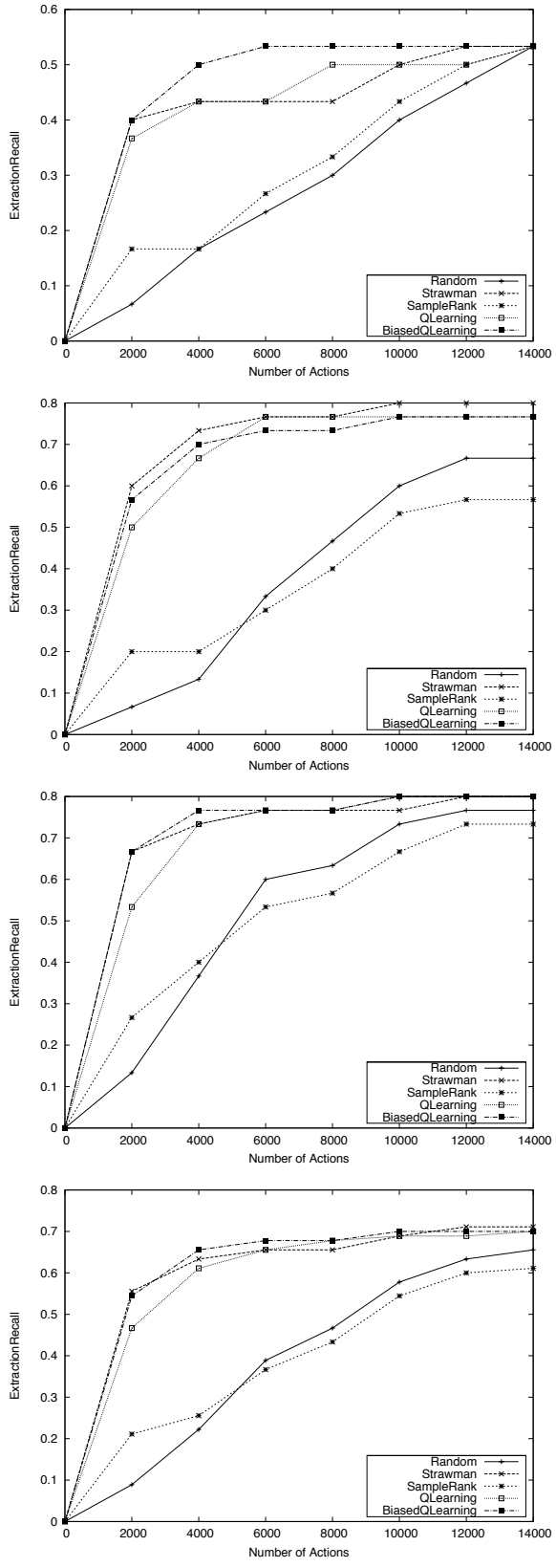

Figure 1: RBIE Using the Oracle. The graphs from top to bottom are : Email, Job Title, Department Name and Total Entries

5.7.2 RBIE Using Extraction Model

We now study the performance of our proposed method using an actual extraction model, M_e. In this case, each action selection strategy needs to balance both precision and recall. We ran 1000 training iterations of SampleRank ($C_n = 1000$, $C_m = 10$, $C_d = 100$, $C_{\bar{d}} = -50$), q-learning ($C_n = 1000$, $C_m = 10$, $C_d = 100$, $C_{\bar{d}} = -50$), and biased-q-learning ($C_n = 0.01$, $C_m = 1$, $C_d = 0.1$, $C_{\bar{d}} = -0.05$) with learning rate, $\alpha = 0.005$, and discount factor, $\gamma = 0.9$ for this experiment. Figure 2 shows the extraction recall, precision, recall and F1 values for total number of entries in the database. In these methods, precision and recall curves go down towards the end of information gathering process due to noise in the web data, and the extraction process. As before, we see that the straw-man method performs better initially. However, its precision and recall drops mid-way through the information acquisition process, and q-learning method performs better. q-learning also comfortably out-performs random and SampleRank approaches. It achieves 92.4% of the final F1, by only using 14.3% of the total actions. This demonstrates the effectiveness of the policy learned by the Q-learner for selecting good actions for information gathering task. We believe that with more accurate labeling and a better extractor, q-learning method can be shown to be even more efficient. Since SampleRank provides a faster learning approach, we would like to further investigate its applicability for this problem domain. A potential improvement on this front could be achieved by defining macro-actions that span multiple actions to better model the delayed reward.

6. CONCLUSION AND FUTURE WORK

In this paper, we formulated the problem of RBIE for the Web as a Markov Decision Process, and proposed the use of temporal difference q-learning to solve it. We learn a policy for effectively selecting information-gathering actions, leading to significant reduction in resource-usage. On our example task of extracting faculty email, job titles and department names, the q-learning based approach is able to achieve 92.4% of the final F1, by only using 14.3% of the total actions. We also compare it to a fast, online, error-driven algorithm called SampleRank [16], and found that q-learning performs better due to it's ability to model delayed reward. We would like to further investigate the use of SampleRank. We also presented a novel extraction technique that can scale well for large scale, information gathering tasks.

The basic formulation of RBIE as an MDP opens up many interesting avenues of research. Use of TD q-learning is one of the first attempts to learn general information gathering policies. More advanced techniques from the reinforcement learning literature, such as SARSA or least-square policy iteration can be explored. This framework is extendable in many ways. We can easily replace the candidate-mention scheme described in the paper with a more sophisticated extraction algorithm. We can also extend the set of information gathering actions defined here to suit the specific needs of a problem, and still use the general MDP framework. Another dimension to explore is the possibility for information acquisition in parallel, which may lead to interesting new methods. The success of such a learning based approach can lead to its application in many resource-conscious, real-world domains.

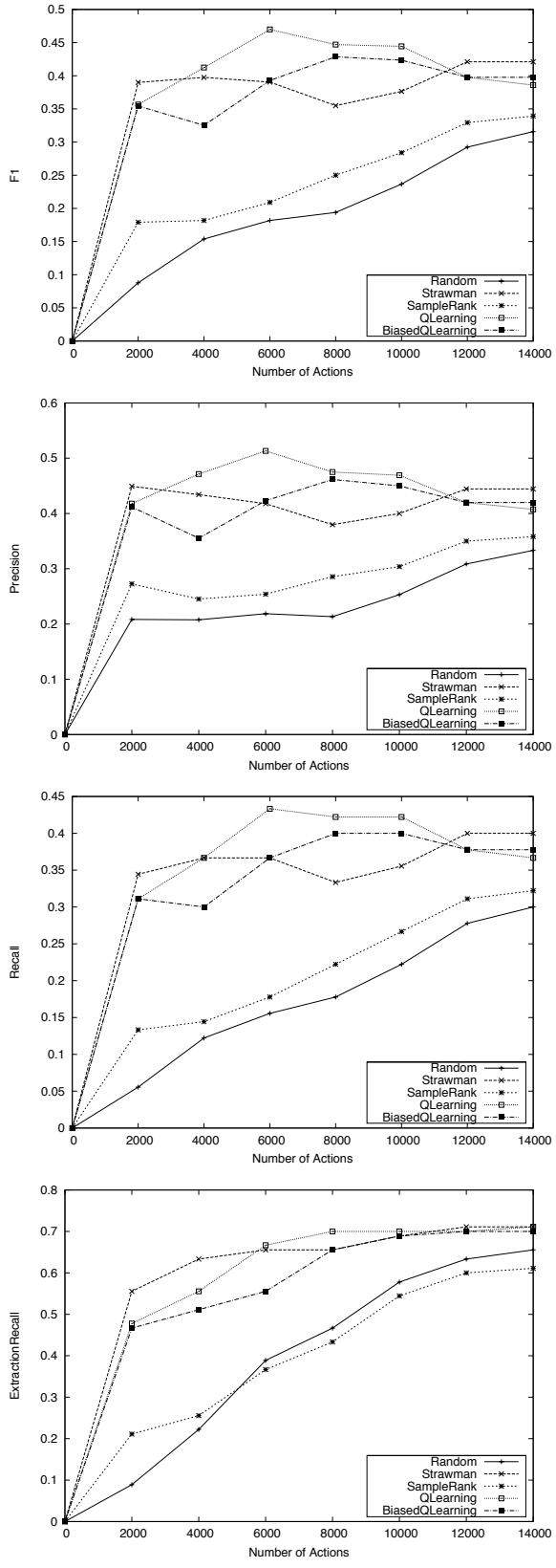

Figure 2: RBIE Using Extraction Model On Total Entries. The graphs from top to bottom are : F1, Precision, Recall and Extraction Recall

7. ACKNOWLEDGMENTS

This research draws on data provided by the University Research Program for Google Search, a service provided by Google to promote a greater common understanding of the web. We thank Michael Wick and anonymous reviewers for useful suggestions. This work was supported in part by the CIIR and in part by the CIA, the NSA, and NSF under NSF grant IIS-0326249 and NSF medium IIS-0803847. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect those of the sponsor.

8. REFERENCES

[1] A. Culotta. *Learning and inference in weighted logic with application to natural language processing*. PhD thesis, University of Massachusetts, 2008.

[2] T. Eliassi-Rad. *Building Intelligent Agents that Learn to Retrieve and Extract Information*. PhD thesis, University of Wisconsin, Madison, 2001.

[3] O. Etzioni, M. Cafarella, D. Downey, S. Kok, A. Popescu, T. Shaked, S. Soderland, D. Weld, and A. Yates. Web-scale information extraction in knowitall. In *WWW*, 2004.

[4] P. Kanani and A. McCallum. Learning to select actions for resource-bounded information extraction. *UMass CS Tech. Report UM-CS-2011-042*, 2011.

[5] P. Kanani, A. McCallum, and S. Hu. Resource-bounded information extraction: Acquiring missing feature values on demand. In *PAKDD*, 2010.

[6] P. Kanani and P. Melville. Prediction-time active feature-value acquisition for customer targeting. In *Workshop on Cost Sensitive Learning, NIPS*, 2008.

[7] C. A. Knoblock. Planning executing, sensing and replanning for information gathering. In *IJCAI*, 1995.

[8] D. J. Lizotte, O. Madani, and R. Greiner. Budgeted learning of naive-bayes classifiers. In *UAI*, 2003.

[9] A. K. McCallum. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu, 2002.

[10] P. Melville, M. Saar-Tsechansky, F. Provost, and R. Mooney. An expected utility approach to active feature-value acquisition. In *ICDM*, 2005.

[11] M. H. Nodine, J. Fowler, T. Ksiezyk, B. Perry, M. Taylor, and A. Unruh. Active information gathering in infosleuth. *IJCIS*, 9(1-2):3–28, 2000.

[12] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach (2nd Edition)*. Prentice Hall, 2003.

[13] V. Sheng and C. Ling. Feature value acquisition in testing: a sequential batch test algorithm. In *ICML'06*.

[14] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[15] C. J. C. H. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, 1992.

[16] M. Wick, K. Rohanimanesh, K. Bellare, A. Culotta, and A. McCallum. Samplerank: Training factor graphs with atomic gradients. In *ICML*, 2011.

[17] F. Wu, R. Hoffmann, and D. S. Weld. Information extraction from wikipedia: moving down the long tail. In *SIGKDD*, 2008.

[18] S. Zilberstein. Resource-bounded reasoning in intelligent systems. *ACM Comput. Surv*, 28, 1996.