Goal-Oriented Knowledge Collection

Yen-Ling Kuo and Jane Yung-jen Hsu
Department of Computer Science and Information Engineering
National Taiwan University
yjhsu@csie.ntu.edu.tw

Abstract
Games with A Purpose (GWAP) has been demonstrated to be efficient in collecting large amount of knowledge from online users, e.g. Verbosity and Virtual Pet game. However, its effectiveness in knowledge base (KB) construction has not been explored in previous research. This paper examines the knowledge collected in the Virtual Pet game and presents an approach to collect more knowledge driven by the existing relations in KB. In this paper, goal-oriented knowledge collection successfully draws 10572 answers for the “food” domain. The answers are verified by online voting to show that 92.07% of them are good sentences and 95.89% of them are new sentences. This result is a significant improvement over the original Virtual Pet game, with 80.58% good sentences and 67.56% weekly new information.

Introduction
Knowledge collection is an essential part in building an AI system. Both domain-specific and common sense knowledge are required for a system to infer new facts or actions. While techniques for mining knowledge from corpus or web pages have been developed (Schubert and Tong 2003; Carlson et al. 2010), it is difficult for computers to discover the common sense knowledge hidden in text (Eslick 2006) or automatically extract all the knowledge in a specific domain.

Since the introduction of Games With A Purpose (GWAP) (von Ahn 2006), it has been used for solving problems that is intractable for computers. Knowledge collection proves to be a good fit with GWAP (von Ahn, Kedia, and Blum 2006; Kuo et al. 2009). As demonstrated in the Virtual Pet game (Kuo et al. 2009), the Chinese common sense was successfully collected and verified via question-answering between players within a year. The collected knowledge has been put into the ConceptNet (Havasi, Speer, and Alonso 2007) knowledge base (KB).

Although GWAP can be used to collect a large amount of data, its ability in drawing informative data has not been fully explored. The analysis in (Weber, Robertson, and Vojnovic 2008) shows that players tend to give easy and generic descriptions in GWAP; therefore, the diversity of the output data is limited. Under the purpose of constructing a KB, this shortcoming diminish the effectiveness of a game to build a complete KB.

In order to fill up the gap between the current common sense KB and the complete common sense KB, this paper examines the amount of new information produced in the Virtual Pet game and presents an approach to populate knowledge in a goal domain (i.e. the featured knowledge domain of a KB.) Knowledge base builders can specify their goals by choosing seed questions for the game. The game will infer new questions from these seed questions to retrieve more related knowledge from users.

This paper starts by analyzing the collected knowledge in the Virtual Pet game, and reviewing related work on knowledge acquisition. Following the experiments on repeated Q&A cycles in the Virtual Pet game, our findings are discussed. We then present the proposed goal-oriented knowledge collection method. An one-week experiment was conducted to show that it helps not only collect knowledge with high precision for a specific goal domain but also increase the amount of new information in each day. This paper concludes with a comparison of the proposed method with the Virtual Pet game without the proposed method.

Virtual Pet Game
The Virtual Pet game is a community-based game that takes the advantage of rich user interactions in a community. (Kuo et al. 2009) The uptodate Chinese common sense knowledge was collected and verified via question-answering between players in the game.

Game Design
In this game, players are motivated to bond with their pets. They can teach their pets common sense in “Common Sense Q&A” to help them get common sense points and become more intelligent. A question in the Virtual Pet game is consisted of a blank, a concept and a pre-defined relation. For example,

- UsedFor(c, b), e.g. Spoon is used for __?
- IsA(b, c), e.g. __ is an animal?

The major interaction takes place in the pet’s home (see figure 1.) The interaction flow of the Virtual Pet game is shown in figure 2. Following are the brief description of each step.
• **Action**: When the players use “Common Sense Q&A”, they may ask/answer their pets questions, do homework (i.e. vote good or bad to an answer), or use the common sense points to exchange food for their pets. These interactions in fact are asking, answering, and voting other players’ questions/answers through their pets. While they do the actions, their pets’ common sense point increased as a reward for them to interact more with their pets.

• **PTT Mail**: After players answer a question or vote an answer, the system will send a PTT mail to the player who provide the question/answer. This PTT mail is shown as if the players’ pet finds the answers in the pet school and wants to show what it learns to its owner.

• **Pet’s school**: Once the players get the answer from their pets, they can go to the pet’s school to reflect the teaching quality, i.e. whether the answer is good or not. If an answer is voted as a bad answer, the system will send a PTT mail to the player to warn him/her not to teach his/her pet this kind of bad knowledge.

**Effectiveness Analysis**

A 75-day data (from August, 2009 to October, 2009) was used for examining the effectiveness of the game. The number of the collected sentences is plotted in figure 3. We can find that at least $\frac{1}{3}$ sentences are already in the KB constructed by the Virtual Pet game. This phenomenon implies that the game’s ability in drawing new data is only $\frac{2}{3}$ than it is supposed to be.

We also discover that the frequencies of answers follow the power law (see figure 4). While most of the answers repeat a small portion of concepts, 1,484,356 concepts are just input once by the players. The concepts that repeat many times are very generic (i.e. easy) rather than specific (i.e. hard) terms in their corresponding domains. As shown in table 1, sentences that contain the generic terms (e.g. “sport” and “food”) are more than the specific terms (e.g. “basketball” and “carrot”).

**Related Work**

Encoding millions of common sense knowledge into machine usable form was time-consuming and expensive. In addition to the Virtual Pet game, there are other approaches for common sense knowledge acquisition. We review some of them in this section.

**Collecting Common Sense from People**

The Cyc project (Lenat 1995) takes 25 years to collect over one million common sense assertions from experts. A team of knowledge engineers carefully encoded knowledge into
logic representation to ensure the correctness and unambiguity of the data. This approach guarantees the highest quality of KB. However, it is expensive, time-consuming, and hard to scale up.

In contrast, the MIT Open Mind Common Sense (OMCS) project (Singh et al. 2002), took the Web.0 approach to harvest common sense from online users. Using the pre-defined templates, over one million sentences have been successfully collected. Nevertheless, this approach still requires a lot of efforts to attract online users to contribute their knowledge on the OMCS website; and thus, the KB grows slowly in non-English extensions.

Motivating Volunteers to Contribute
Games With A Purpose (GWAP) (von Ahn 2006) approach utilizes computer games to gather players and guide them to perform tasks while playing the game. It has demonstrated its usefulness in various domains, such as image labeling (von Ahn and Dabish 2004) and common sense collection (von Ahn, Kedia, and Blum 2006). Compared to the Web.0 approach, knowledge base builders have more power to push users to contribute specific knowledge via the mechanism of the game.

As the Virtual Pet game, Verbosity (von Ahn, Kedia, and Blum 2006) and Common consensus (Lieberman, Smith, and Teeters 2007) are games for common sense collection. Verbosity is two-player game in which the Narrator gives clues to help the Guesser figure out a secret word. The clues are collected as common sense knowledge about the corresponding secret word. Common consensus is a game to collect human goals and validate them automatically.

Although this kind of games successfully motivated players to contribute common sense assertions, their contribution to the construction of KB has not been evaluated. In this paper, we use the Virtual Pet game as our experimental platform to push questions to players and explore the effectiveness issue of this approach.

Repeated Q&A Cycles
In order to reduce the redundant answers (see figure 3) in GWAP, we analysed the process of asking a question many times to figure out how the redundant data produced. Before illustrating the experiments, we introduce some terminologies and our hypothesis.

Notation and Hypotheses
A Q&A cycle is consisted of a question and its corresponding answers. If a question is asked many times, the same Q&A cycle repeats. To evaluate the new information produced by each question, we use $C_i$ to denote the answers to a specific question generated at $ith$ day from its start and define the amount of new information produced at $kth$ day as follows:

$$NewInfo_{k-day} = (1 - \frac{\bigcup_{i=1}^{k-1} (C_i \cap C_k)}{|C_k|}) \times 100\% \quad (1)$$

Two hypotheses on the redundant answers are examined in the section.

- **Hypothesis 1**: The amount of new information decays over time since people in the same group share similar knowledge.
- **Hypothesis 2**: The rate of decay depends on the size of domain knowledge.

Experimental Setup
We used spreading activation (Collins and Loftus 1975) on Chinese ConceptNet to estimate the size of domains. For each domain, we find its related concepts by traversing from the name of the domain and spreading outwardly to include other concepts. The estimated domain size is the number of the related concepts versus the number of concepts in the KB. Three domains of different sizes were chosen for this experiment; they were “electric appliances”, “food”, and “sport” (see table 2). Questions were created for each domain (see table 3) and repeatedly pushed to different players in the Virtual Pet game.

<table>
<thead>
<tr>
<th>Generic terms</th>
<th># of sentences collected</th>
<th>Specific terms</th>
<th># of sentences collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sport</td>
<td>5,073</td>
<td>Basketball</td>
<td>341</td>
</tr>
<tr>
<td>Food</td>
<td>2,157</td>
<td>Carrot</td>
<td>142</td>
</tr>
</tbody>
</table>

Table 1: Example of generic and specific terms

Table 3: Sample questions in the experimental Q&A cycles

<table>
<thead>
<tr>
<th>Electric appliances</th>
</tr>
</thead>
<tbody>
<tr>
<td>The appliance you will use when you <strong>sleep</strong> is __?</td>
</tr>
<tr>
<td>The appliance you will use when you <strong>watch TV</strong> is __?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Food</th>
</tr>
</thead>
<tbody>
<tr>
<td>_ is a kind of <strong>noodle</strong>?</td>
</tr>
<tr>
<td>_ is a kind of <strong>sauce</strong>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sport</th>
</tr>
</thead>
<tbody>
<tr>
<td>The thing you will do when you <strong>play volleyball</strong> is __?</td>
</tr>
<tr>
<td>The thing you will do when you <strong>play soccer</strong> is __?</td>
</tr>
</tbody>
</table>

Experimental Result
We collected answers to the questions listed in table 3 for 9 days and acquired 50, 200, and 100 answers everyday for questions about electronic appliances, food, and sport. The amount of new information produced by a question in $kth$ day is calculated and shown in figure 5. In figure 5, the amount of new information decreases with time for all questions. The lines even approach zero on 7th and 8th day in “electronic appliances” and “food.” These results confirm

1http://openmind.media.mit.edu/
Table 2: Estimated domain size

<table>
<thead>
<tr>
<th>Domain</th>
<th>Electronic appliances</th>
<th>Food</th>
<th>Sport</th>
</tr>
</thead>
<tbody>
<tr>
<td># of related concepts</td>
<td>476</td>
<td>809</td>
<td>1245</td>
</tr>
<tr>
<td>Estimated domain size</td>
<td>$0.00738 \times</td>
<td>KB</td>
<td>0.0125 \times</td>
</tr>
</tbody>
</table>

$KB$ is concepts in the knowledge base $|KB| = 64450$ in Chinese ConceptNet

Figure 5: Amount of new information produced by different questions per day

hypothesis 1 and can be interpreted from the following three aspects.

- **Gaming community**: The players in the same community share similar knowledge. If a question is asked many times, the players will explore all the knowledge they have. Hence, the knowledge we explored later is with higher probability of being explored by other players and add little information to the KB.

- **Domain knowledge**: Even if the common sense knowledge changes over time, the domain knowledge remains the same and is finite at a specified time. It is possible to enumerate all the knowledge with regard to a certain domain. In the repeated Q&A cycles, players behave like knowledge engineers and codify knowledge by enumerating them. The amount of new information approaches zero once the players enumerate all the instances.

- **Gaming strategy**: Since the game awards points to players who provide the right answers without considering the informativeness of answers, the players tend to answer generic terms, e.g. “people can eat vegetables”, rather than specific terms, e.g. “people can eat cabbage.” This tendency eases tasks for players but leads to the overlapped answers in 3.

The equations in figure 5 show the results of nonlinear regression analysis for the three domains. The coefficients of $\ln(x)$ are smaller than -0.42 for electronic appliances ($R^2 > 0.8$), around -0.39 for food ($R^2 > 0.7$), and -0.35 for sport ($R^2 > 0.95$). These results confirm hypothesis 2 and reveal that the larger the domain knowledge is the slower its new information decays while a question is asked many times.

**Goal-oriented Knowledge Collection**

The common sense KB is consisted of many domain knowledge. In order to fill up the gaps between the current KB and the complete KB, we need to collect knowledge for each domain respectively. In this section, we introduce goal-oriented knowledge collection using human computation game to collect knowledge for different domains. Here, the “goal” refers to the knowledge domain we’d like to populate in a KB.

**Relation Network**

Relation network is used for telling the game how to infer new questions. Since a concept can serve as an intermediate concept of two relations, we define relation network to build the implicit connection between different relations via the intermediate concept. The node is a relation of a question and the link is defined as “if $\text{Relation}_1(x, y)$ and $\text{Relation}_2(y, z)$ are true for every possible $x, y, and z$, there is a link between $\text{Relation}_1$ and $\text{Relation}_2$ in the relation network.” Hence, the game can infer new questions by traversing in the relation network. For example, if we get an answer “classroom” to the question
“You are likely to find _ in a school?”, we can traverse from AtLocation to HasSubevent to ask players a new question “One thing you will do when you in classroom is _ ?”

The relation network in the Virtual Pet game is shown in figure 6 where the relations are from the ConceptNet. Games with different purposes can build their own relation network based on the concepts and relations it would like to acquire. For a game that aims for collecting “events”, we can choose event-related ones such as HasSubevent or Causes.

**Game as Algorithm**

In goal-oriented knowledge collection, we treat the Virtual Pet game as a subroutine we can call. Algorithm 1 provides the detailed VirtualPet procedure for collecting answers to a set of questions $Q$. If an answer is good enough, it will be put into the $KB$. A threshold $\theta$ is set to avoid asking a question too many times. In the current implementation of the Virtual Pet game, the criteria of good answer is the answer count $> 1$ or good votes $> \text{bad votes}$.

**Algorithm 1 VirtualPet($KB, Q, \theta$)**

Require: A knowledge base $KB$, question set $Q$, and a threshold $\theta$ of $\text{NewInfo}_{\text{day}}$

Ensure: A set of common sense sentences $CS$

1: for all $q \in Q$ do
2: \{Ask $q$ until its information is enough\}
3: repeat
4: $CS' = \text{Ask}(q)$
5: for all $cs \in CS'$ do
6: \{Put the sentence to $KB$ if it is good\}
7: if (answer count of $cs > 1$) or (good votes of $cs > \text{bad votes of } cs$) then
8: $CS \leftarrow CS \cup cs$ and $KB \leftarrow KB \cup cs$
9: end if
10: end for
11: Calculate $\text{NewInfo}_{\text{day}}$
12: until $\% \text{ of new information} < \theta$
13: end for
14: return $CS$

**Question Expansion**

Setting a threshold of $\text{NewInfo}_{\text{day}}$ in VirtualPet only prevents the overlappings of answers. For the purpose of guiding the knowledge collection in a goal domain, the game should equip the ability to push new questions to players actively. Algorithm 2 shows the recursive QuestionExpansion procedure for inferring new questions from answers, where a relation network is required and an expansion depth $d$ is set to avoid infinite repetition. QuestionExpansion calls VirtualPet every time it is evoked and uses the results returned by VirtualPet to create new questions via traversing the relation network.

**Algorithm 2 QuestionExpansion($KB, Q, RN, \theta, d$)**

Require: A knowledge base $KB$, seed question set $Q$, relation network $RN$, threshold $\theta$ of $\text{NewInfo}_{\text{day}}$, and the expansion depth $d$.

Ensure: A set of common sense sentences $CS$

1: if $d = 0$ then
2: $CS \leftarrow CS \cup \text{VirtualPet}(KB, Q, \theta)$
3: else
4: $CS' \leftarrow \text{VirtualPet}(KB, Q, \theta)$
5: $CS \leftarrow CS \cup CS'$
6: \{Create new questions\}
7: for all $cs \in CS'$ do
8: relation $\leftarrow$ next relation of $cs$ in $RN$
9: new question $\leftarrow$ concept in $cs +$ relation
10: $Q' \leftarrow Q' \cup$ new question
11: end for
12: $CS \leftarrow CS \cup$
13: QuestionExpansion($KB, Q', RN, \theta, d - 1$)
14: end if
15: return $CS$

**Evaluation**

An one-week experiment was conducted to see whether the proposed procedure improves the informativeness of answers. “Food” was chosen as the goal domain and five seed questions were selected; they were “_ is a kind of noodle / sauce / dessert / drink / vegetable ?”. The parameters in QuestionExpansion procedure were decided empirically. $KB$ was the Chinese ConceptNet, $Q$ was the seed questions, $RN$ was defined in figure 6, $\theta$ was set to 75%, and $d$ was 10 in our experiment.

**Data Quality**

755 questions were created in one week, and 12 of them were reported as bad ones. However, 1/3 of the bad questions were reported because of typo or part-of-speech error. This record shows that relation network is useful for inferring new questions.

In addition, 10,572 sentences were collected within a week, and 9,734 of them were voted as good common sense by the Virtual Pet game players, i.e. the precision is 92.07%. Compared with the game without QuestionExpansion procedure (80.58% good common sense), we think the goal-oriented knowledge collection approach is practical, and the
questions inferred by relation network can guide players to give more good answers.

New Information
Since we change question once its \( NewInfo_{k-week} \) is below a threshold \( \theta \), we modified Eq. (1) to compare sentences produced in each week. We use \( C_i \) to denote the answers generated in \( i-th \) week from the start of the game and compute the amount of new information produced at \( k-th \) week as follows:

\[
NewInfo_{k-week} = (1 - \frac{\sum_{i=1}^{k-1} (C_i \cap C_k)}{|C_k|}) \times 100\%
\]

\( NewInfo_{k-week} \) in our experimental period is 95.89%. It is considerably higher than the Virtual Pet game without QuestionExpansion (about 67.56% for each week). We think the threshold \( \theta \) set in VirtualPet plays an important role so that players are not likely to answer the same questions many times.

Coverage

<table>
<thead>
<tr>
<th>Inference chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>pasta isA noodle</td>
</tr>
<tr>
<td>pasta MadeOf flour</td>
</tr>
<tr>
<td>flour AtLocation supermarket</td>
</tr>
<tr>
<td>supermarket HasA noodle</td>
</tr>
<tr>
<td>beef noodle isA noodle</td>
</tr>
<tr>
<td>beef noodle AtLocation noodle shop</td>
</tr>
</tbody>
</table>

The example of inferred questions is shown in table 4, where the italic concepts are answers to the questions. From the example, we know that more details of a concept are provided by players in the inference. The coverage of a concept can be calculated by the number of concepts involved in the spreading activation from that concept. The coverage of “dessert” in our experiment is 1763, which is higher than 975 in the Virtual Pet game without QuestionExpansion procedure. We think the high coverage is better for the construction of complete common sense KB.

Conclusion
This paper proposed an approach to collect knowledge for a goal domain. By treating the game as a learning algorithm that can adapt to the knowledge players contribute, we are able to collect more good knowledge. Experiment has been conducted in the “food” domain. With only 5 seed questions, the proposed method succeeded in generating 743 good questions and 9734 good sentences with over 90% new information. In order to diversify the goals in the Virtual Pet game, we released a service on the website\(^2\). Users can specify their goals on the website for the Virtual Pet game to collect. Thereby, it is possible to generate knowledge with new information automatically for Q&A based human computation games.

\(^2\)http://agent.csie.ntu.edu.tw/commonsense/

References


