ENHANCING IMAGE Annotation games by using Asian WordNet

Virach Sornlertlamvanich and Thatsanee Charoenporn

National Electronics and Computer Technology Center, Thailand
virach.sornlertlamvanich@nectec.or.th, thatsanee.charoenporn@nectec.or.th

ABSTRACT

ESP and Peekaboom games are successfully done to engage annotators in labelling the images and locating the objects in the images. The advantage of these games is making annotators feel like enjoying a game rather than working hard to label the ton of images. We extended Peekaboom game by introducing Asian WordNet (AWN) to expand the guessed label in terms of its synonym, hypernym, hyponym, meronym, and holonym. The result yields in a higher matched ratio. Moreover, AWN will also provide the semantic link between 13 Asian languages and the English language. Our method also shows a potential in creating cross-language image labelling game.

Keywords: ESP game, Peekaboom game, Image labeling, Asian WordNet

1. INTRODUCTION

von Ahn, L., et al., 2004 [1] proposed a game approach to associate word to a image called ESP game. ESP game is used to match the name of the image proposed by two players. The name of the image is determined when the two players guess the same name without knowing each other’s guess. The result is assumed to confirm the name proposed by at least two parties. Once the image is named, Peekaboom game [2] is extendedly proposed to locate the object in the image. Image name and the object appeared in the image may not always be the same due to where the player is interested in when looking at the image. Given an image and the name of the image, Player one gradually opens the visible area to Player two by selecting the area where it can provide the critical area to recognize the name of the object. Peekaboom also provides a mechanism for Player one to give a hint about how the word relates to the image, i.e. noun, related noun, text, and verb. Players can enjoy the game and have fun in guessing words when looking at an image, while the object in the image is located and labeling precisely. These labeling images can be used in as many aspects as we have to deal with the rich information of images such as image search, image mapping, image understanding and so on. At least, the current approaches still need a huge amount of image data to train for a model in machine learning paradigm.

Cultural database is a collection of national culture, which composed of a media with the description text. The media can be any type of image, video, flash animation, PDF, or 360-degrees panoramas. The contents are surveyed and collected by local people who are familiar with their living. Before releasing the content, the authorized members from the Ministry of Culture have to examine the quality and the degree of information. These are all made available on the social networking system at http://www.m-culture.in.th/ [6]. Since January 2011 until present (October 2012), there are 78,028 titles and 273,983 medias uploaded.

We are going to label the object in the image from the above cultural database. Image title and tags are used to create the choices for matching in ESP-like game when it is a single player game. Otherwise, the image name will be determined when both players have the same guess.

From the result of ESP-like game, we propose the extended Peekaboom game by using AWN [3,4]. The label from the ESP-like game is expanded in terms of its synonym, hypernym, hyponym, meronym, and holonym. The terms in the sets of related terms can help matching the guess from the player. By the nature of naming the image, the label can usually be a specific object in the image, or a component of an object, or whole group of an object. An object can also be called in several ways. In this case, the set of synonym can help matching the guess word. The degree of generality in calling is also another concern. In this case, the set of hypernym and hyponym can cover the possible recognition of the player. Meronym and holonym can also extend the possibility of matching words according to the focus of the player on the object. It can be a part of the object, or the object as a whole.

The based data of the cultural database are created in the Thai language. However, the recognition of the image is language independent. Once we can convert the original label in to other languages, the game will be enjoyed in the cross language manner. The related terms in the synset from 13 languages including the English language of AWN are prepared for the guess word matching. From AWN, first the synset ID of the original Thai word is extracted, and then the synsets of other languages are obtained by using the extracted synset ID.
As a result, the extended Peekaboom game can raise the image-word pairing ratio and realize the cross-language image labeling.

2. ESP-LIKE GAME

To prepare image-word pairing data, we deploy some rules to provide word hint to the player of ESP-like game in case of single player mode. The image title and tags are available from the cultural database. These words usually show a kind of relation to the image. The title can be the name of an object, or an event shown in the image. Compound nouns are very normal in Thai, so that we need a word segmentation tool [5] to list up the component words. The tags can be the other possible calling names, or the generic names.

Figure 1 Japanese bicycle home

Figure 1 shows a record of บ้านรถยนต์ญี่ปุ่น (Japanese bicycle home) as the title and บ้านรถปั่น (bicycle home) as the tag. In this case, the tag “bicycle home” is the generic or common name of the title “Japanese bicycle home”.

In two players mode, the player has a trend to guess the generic word, and give a name to a specific object rather than the image as a whole. In the same example of Figure 1, both players have a trend to guess the successful match on “bicycle” rather than “bicycle shop”.

From this ESP-like game, we successfully named the images or at least obtained a list of candidates for labeling the object in the image to be used in the next extended Peekaboom game.

3. EXTENDED PEEKABOOM GAME

von Ahn, L., et al., 2004 [1] has successfully done in proposing a kind of online game to draw attention from collaborators in annotating images. ESP game is used to confirm on the name of image title. Upon providing an image, if more than one player agrees to call with the same name then the image should be called by that name. The game is played online and each player is not allowed to see each other’s guess. The name of the image is settled when the names typed in by both players are matched. Peekaboom game [2] is the extension of the ESP game. It is for labeling the object in the image. In the original game, Player one is provided an image and the title. Player one opens the area in the image, which is called by the title. This is to show Player two the image area and wait for the guess from Player two. Player two will guess until it matches the image title. The result will refine the name of the object shown in the image. The task is quite tough when the image title is usually precisely defined (by an original creator) but not the person who guesses. For example, Figure 1 is more likely to be guessed as a “shark” rather than a “leopard shark”.

The extended Peekaboom game is proposed to make the guess word matched with the image title in a more efficient way. AWN is introduced to generate a set of words from expanding the image title in terms of its synset, hypernym, hyponym, meronym and holonym. The terms in the sets of related terms can help matching the guess from the player. By the nature of naming the image, the label can usually be a specific object in the image, or a component of an object, or whole group of an object. An object can also be called in several ways. In this case, the set of synonym can help matching the guess word. The degree of generality in calling is also another concern. In this case, the set of hypernym and hyponym can cover the possible recognition of the player. Meronym and holonym can also extend the possibility of matching words according to the focus of the player on the object. It can be a part of the object, or the object as a whole.

3.1 Playing the extended Peekaboom game

Figure 3 Screen seen by Player One and Player Two

Player one selects an area to display to Player two by using a rectangle selection frame. Player two guesses a proper word according to what is being seen with in a
limited time. In Figure 3, screen a) is seen by Player one and screen b) is seen by Player two.

Figure 4 shows the successful guess “จักรยาน” (bicycle) for the object. The result also shows the matched level of exact, synset, hypernym, and hyponym. This means that in addition to matching the image title exactly, the word “จักรยาน” is commonly appeared in the sets of synonym, hypernym, and hyponym.

In case that Player two guesses the more generic name as shown in Figure 5, only a hypernym “พาหนะที่มีล้อ” (wheeled vehicle) can be matched with the guess word.

Table 1 Related synsets of the word “จักรยาน” (bicycle)

<table>
<thead>
<tr>
<th>Synset</th>
<th>Hyponym</th>
<th>Meronym</th>
</tr>
</thead>
<tbody>
<tr>
<td>จักรยาน</td>
<td>จักรยานเสือภูเขา</td>
<td>รถจักรยาน</td>
</tr>
<tr>
<td>รถจักรยาน</td>
<td>รถจักรยานเสือภูเขา</td>
<td>รถจักรยาน</td>
</tr>
<tr>
<td>เล็กที่สุด</td>
<td>จักรยานเสือภูเขา</td>
<td>รถจักรยาน</td>
</tr>
<tr>
<td>จักรยาน</td>
<td>จักรยานเสือภูเขา</td>
<td>รถจักรยาน</td>
</tr>
<tr>
<td>รถจักรยาน</td>
<td>รถจักรยาน</td>
<td>รถจักรยาน</td>
</tr>
<tr>
<td>รถจักรยาน</td>
<td>รถจักรยาน</td>
<td>รถจักรยาน</td>
</tr>
<tr>
<td>รถจักรยาน</td>
<td>รถจักรยาน</td>
<td>รถจักรยาน</td>
</tr>
<tr>
<td>รถจักรยาน</td>
<td>รถจักรยาน</td>
<td>รถจักรยาน</td>
</tr>
<tr>
<td>รถจักรยาน</td>
<td>รถจักรยาน</td>
<td>รถจักรยาน</td>
</tr>
</tbody>
</table>

In all cases, using AWN for keyword expansion can help image-word pairing in an efficient way. This is because it is more general that people can think about

the generic term rather than the specific one. In addition, the variation of naming should also be allowed to call the same object. Therefore, synset from AWN is quite useful to produce the successful image-word pairing.

3.2 Experiment result

We tried on 18 images played by 19 persons. For each image, we allow 60 seconds to guess a proper word. Next image will appear after the player can guess for the matched word. The players can pass the image if they found that they could not find a proper word in case.

Table 2 Number of successful matched word

<table>
<thead>
<tr>
<th>Exact</th>
<th>Syn</th>
<th>Hyper</th>
<th>Hypo</th>
<th>Mero</th>
<th>Holo</th>
</tr>
</thead>
<tbody>
<tr>
<td>229</td>
<td>32</td>
<td>16</td>
<td>1</td>
<td>7</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 2 shows the number of matched word in each case. Synonym can additionally provide a match in 32 cases, hypernym in 16 cases, hyponym in 1 case, meronym in 7 cases, and holonym in 11 cases. This confirms that generic terms provided by hypernym have more chances in matching than specific terms provided by hyponym. In case of the matched holonym, the holonym of “พิภนLIGHT” (stalagmite), “ถ้ำ” (cave) is suggested for the reason of familiarity. The word “ถ้ำ” (cave) has a more frequently use than the special term like “พิภนLIGHT” (stalagmite).

In total, AWN can expand the matching in 67 cases or increase 22% of matching ratio. All cases of relation expansion in AWN, i.e. synonym, hypernym, hyponym, meronym, and holonym have a significant contribution in expanding the chance in image-word pairing.

4. CROSS LANGUAGE PEEKABOOM GAME

The image title is expended by looking up the synset of other languages in AWN. In AWN, there are 13 Asian languages available and having synset ID linked to the English WordNet (PWN, Princeton WordNet) [7]. The image titles in other languages available in AWN are prepared beforehand by looking up the synset ID for the original Thai image titles. Once the synset IDs are identified, we use the synset IDs to extract the related synsets in other languages to create the possible set of words in each language. As a result, a player can guess a word in any of the available languages. Some possible words in the synset will allow more possible chance to match with the guess word. We then obtain the list of the image labels in other languages.

Figure 6 shows the successful cases in image-word pairing in both English and Japanese. In the same example of “จักรยาน” (bicycle), the player guesses “bicycle” in English and “จักรยาน” in Japanese. Both are the words extracted from the synset of AWN.
5. CONCLUSION

ESP and Peekaboom games can efficiently be used to obtain the labeled image and the object in the image. Players do not feel so big burden in providing the image label. Instead, they enjoy the word guessing and have fun when obtaining the correct label with the score. Guessing a word from different viewpoint can have difficulty in finding the exact words matched to either the words from other player or the image title. Due to the tendency in naming in the general term of the image title and the variety in calling the object, we therefore proposed the term expansion method by using AWN. The structure of AWN can provide the set of synonym, hypernym, hyponym, meronym, and holonym. As a result, we can increase about 22% of matching ratio. In addition, AWN has also a synset ID linked to other 13 Asian languages and the English language. This made it possible to label the image in a cross-language manner.

6. ACKNOWLEDGMENTS

We thank Wirat Chinnan and Thanasan Tanhermhong for their hard work in preparing the system and experiment environment. We also thank the volunteers from NECTEC in playing the games in the test environment. Without the huge data from the cultural database development project of the Ministry of Culture, we do not think that we can realize the system.

7. REFERENCES