Labeling Images with a Computer Game

Luis von Ahn and Laura Dabbish

School of Computing Science, CMU

Presented By: Akshay Gadi Patil

Google Image Search Results

For Car (2003-2004)

Source: Google image search 2004
(according to this paper)

For Car (2016)

Source: Google Image Search
(3rd Oct 2016)

Why….. How…..

- Text based image retrieval annotate images with the text derived from HTML documents displaying them.

- Text can include:

  Image Caption, Text surrounding the image, filename of HTML document, entire text in the page, etc.

- Not the right approach.

Motivation

- What is the problem with the existing image search ? (back in 2004)

  The text adjacent to images is often scarce, and can be misleading and hard to process*.

- Can every image available on the web have some label associated with it ?
- If yes, can the labeling be made accurate(rather, more accurate) ? May be by crowdsourcing!

*Carson, Chad, and Virginia E. Ogle. "Storage and retrieval of feature data for a very large online image collection." *IEEE Data Eng. Bull.* 19.4 (1996): 19-27.

Outline

- Introduction
- “The ESP Game”
- Conclusion

Introduction

- Accurate descriptions of images are required by many applications

  Ex: image search engines, accessibility programs for the visually impaired.

- The only method available for obtaining precise image description was (/is?) manual labeling – tedious and costly.

- What if this task of labeling the images is made enjoyable, without people realizing it!

Introduction (contd.)

- A new interactive system is presented in the form **a game** for image labeling tasks.
- The people who play the game, label the images.
- Significant contributions-

  1) the way this work addresses the image labeling problem,
  2) makes use of people's existing perceptual abilities rather than computer vision techniques.

Why not Computer Vision and Machine Learning techniques ?

- Absence of large databases of labeled images needed as training sets for machine learning algorithms. (Ex: ImageNet*)

- Existing (in 2004) computer vision techniques didn't produce a solution that could determine the contents of images in a widely useful way.

* Deng, Jia, *et al.* "Imagenet: A large-scale hierarchical image database."*Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.

“The ESP Game” – General description

- It’s an online game.
- 2 players in one game; multiple parallel sessions can occur.
- Each person can be in one game at one time.
- Partners randomly assigned.

"The ESP Game" – General description (contd.)

- Players are not told who their partners are.
- What do they have in common? **Only** the Image
- Goal is to predict what the other player is typing., think like one another

  (**E**xtra **S**ensory **P**erception)

“The ESP Game” – General description (contd.)

Player 1 guesses: purse
Player 1 guesses: bag
Player 1 guesses: brown

Success! Agreement on “purse”

Player 2 guesses: handbag

Player 2 guesses: purse
Success! Agreement on “purse”

Guesses can only be 13 characters long

“The ESP Game” – General description (contd.)

- Next image appears when both have typed the same string (“agreeing on an image”).
- Not necessarily at the same time but at some point of time while the image is on the screen.
- One game duration - 2.5 minutes
- Max number of images per game- 15

“The ESP game”


The ESP Game - Netscape 6
0:11
Time Left
The ESP Game
2100
score
Taboo Words
MAN
BEARD
Your Guesses
HAT
Type your next guess:
Pass

Details of the game

- Label Threshold:

  Used for assigning a label to an image and for categorising a label as a taboo word

- A word becomes a label (and a taboo word) when ‘X’ number of pairs agree on it.

- The label threshold would then be ‘X’.

Details of the game(contd.)

- Taboo Words:
  - Associated with every image
  - Can't be entered as guesses(Singulars, plurals or phrases containing the taboo word cannot be used)
  - Obtained from the game
- Ex: Second occurrence of an image across games => First taboo word

  => Previously agreed word on that image
- 6 Taboo word in this implementation of the game.

Details of the game(contd.)

- Passing an Image: (When it's no longer enjoyable)
  - When the image is too complex to guess
  - Or when the image has acquired an extensive list of taboo words

- Repeated passing => Image should no longer be used in the game
- Fully labeled images can be re-inserted at a later point of time(after several months or years) because that image may then be used in a different context.

Details of the game(contd.)

- The game is implemented as a Java applet.
- Responsible for running the game and storing information.
- Game server starts a game every 30 seconds (at least one player must be logged in).

“bot”: Pre-recorded game play

- A player can be paired against a “bot”, which is the pre-recorded set of actions from an earlier game session involving two people.
- It doesn’t stop the labelling process.
  - Agree on a new word, we get a new label.
  - Agree on a pre-recorded label, confidence of that label increases.
- Useful when the game is still gaining popularity as the crowd knowing about the game may not be large.

How is cheating handled?

- Different ways of cheating:
  - Communicate with the other player (Sol: Random pairing in distributed locations reduces this probability).
  - Getting paired with themselves (Sol: Pairing happens only if IP addresses are different)
  - Large group of players agreeing on a unified strategy, ex: agree to type ‘a’ on every image (Sol: Inserting a large number of bots when massive agreement strategy is detected).

Image Selection

- Initially, 350,000 images were chosen at random.
- Re-introduce images into the game several times until they are fully labeled.
- Criteria for selecting the images:
  - No blank images, images with a single color, images smaller than 20 pixels on either dimension, images with aspect ratio greater than 4.5 or smaller than 1/(4.5).

Labels from the game

Dog

Leash

German

Shepard

Standing

Canine

Evaluation Studies

- Quality of labels – Search Precision
- Experimental comparison
- Subjective evaluation of labels obtained from the game.

Evaluation Studies

- Quality of labels – Search Precision
- Experimental comparison
- Subjective evaluation of labels obtained from the game.

- Examine the results of searching for all images associated to particular labels.
- 10 labels were chosen at random from the set of all images obtained from the game.
- Choose from labels, that occur in more than 8 images.

Image search result for the label query “Car”

Evaluation Studies

- Quality of labels – Search Precision
- Experimental comparison
- Subjective evaluation of labels obtained from the game

Experiment

- 15 participants who never played the game.
- 20 images were shown at random out of the first 1023 images having more than 5 labels.
- Asked to type six individual words (< 13 characters) best describing the contents in the image.

Results

- For all the 20 images, at least 5 of the six labels produced by the game were covered by the participants ,i.e., each of these labels was entered by at least one participant.
- 3 most common words entered by the participants were present among the labels in the game.

Evaluation Studies

- Quality of labels – Search Precision
- Experimental comparison
- Subjective evaluation of labels obtained from the game.

Experiment

- 15 participants who never played the game.
- 20 images with their corresponding labels (six) from the game were shown at random out of the first 1023 images having more than 5 labels.
- 2 questions:
  - How many of these labels would you use to describe the image to a person who couldn't see?
  - How many of the words have nothing to do with the image?

Results

- For the 1st question, mean was 5.105 words, i.e., 85% of labels corresponding to one image would be useful in describing it.
- For the 2nd question, mean was 0.105, i.e., 1.7% of labels corresponding to one image were thought to have nothing to do with the image.

Extensions to “the ESP game”

- Content Specific labeling
  - Use of theme rooms (Ex: Painting, Sports etc.)
  - Image within the theme rooms are domain specific and players who wish to have specific domain of images, can play them.
  - Labels from such a game are likely to be more specific.

- These theme rooms can be prepared using web directories or using labels obtained from the “general category” ESP game.

Extensions to “the ESP game” (contd.)

- Inappropriate content filtering
  - Small percentage of images on the web are inappropriate for children. So the “general category” ESP game may also be inappropriate for children.
  - Children’s version of the game: (images with certain number of labels + text based filtering) would prevent inappropriate images from reaching children.

Strengths of “the ESP game”

- Creative approach to a hard problem
- Fun to play
- Vast majority of labels are appropriate
- Powerful idea: Reaching consensus with little or no communication

Shortcomings of “the ESP game”

- Finds mostly general labels.
- Lot of redundancy in the labels*.
- Already many images with such labels.

* Weber, Ingmar, Stephen Robertson, and Milan Vojnovic. "Rethinking the ESP game." *Proc. of 27th intl. conf. on Human factors in Computing Systems, ser. CHI*. Vol. 9. 2008.

Conclusion

- Proper labels associated to each image on the Web could allow for more accurate image retrieval.
- Instead of developing a complicated algorithm, the work presents a crowdsourcing technique in the form of a game, making a tedious task enjoyable.
- The game could also be used for other multimedia applications (however, the success of it depends upon how enjoyable the game is for that particular application.)

Thank You!