Project Proposal

Exploring Domain Adaptation for the Image Description Task

Glebys Gonzalez, Maria L. Pacheco

School of Industrial Engineering, Department of Computer Science

`gonza337@purdue.edu, pachecog@purdue.edu`

October 2, 2015

1   Introduction

What is the project

Motivation

2   Related work

Describe any previous work you found related to your project

3   Problem formulation

Describe your project as a machine learning problem, identify inputs objects, labels, possible features

The problem of text based image description has been formulated as a ranking task . Taking the set of images and their descriptions, we can evaluate for each description how well the proposed system ranks an image in contrast to all other images in the set. This framework allows for the analogous task to be evaluated as well, testing how well the system ranks the captions of a given image against the complete caption set.

For both tasks in this cross-modal evaluation the input objects are pairs of the form (i, C) where i corresponds to an image and C corresponds to a set of sentences describing such image. Depending on the task, the pairing would be done from image to captions or from captions to images.

Metrics can be computed automatically, measuring the recall at different levels. For example, the rate of queries in which the correct response was among the top k results and the median rank of the correct responses **?**.

Different approaches for this task have tried different feature representations . We intend to evaluate them further to propose an appropriate feature representation for this project.

Microsoft Research released a dataset of abstract scenes and descriptions following the intuition that extracting high-level semantic meaning from real images tends to be difficult. Real images possess a great amount of detail and complex features that are not necessarily needed to capture its semantic meaning.

We propose to explore domain adaptation in this scenario, having two sets: the abstract dataset as the source set and a repository of real images as the target set. The main goal is to learn the correlation of images and sentences in this simplified domain and map it to the more expressive domain of real images.

4   Data and Evaluation plan

As a domain adaptation problem, we need to define our source and target datasets.

- **Source dataset:** The Abstract Scenes Dataset released by Microsoft Research (Clipart). The first version of this dataset contains 10,000 abstract images of kids playing in a park. The images are composed by a set of 80 pieces of clipart representing 58 different objects. These images are arranged in groups of 10, displaying different pictorial versions of the same scene text description. A newer version of the dataset has been released expanding to 6 the number of sentences associated with each image, amounting for a total of 60,000 sentences .

- **Target dataset (op1):** The Flickr30k Entities dataset is an extension of the Flick30k released by University of Illinois at Urbana Champaign. This dataset contains 30,000 images and 158,000 captions. The dataset has been expanded with an additional 244,000 coreference chains linking mentioned entities in the captions the segment of the images that frame those entities .

- **Target dataset (op2)**: COCO is a dataset released by Microsoft Research containing more than 300,000 images and 5 captions per image. This dataset also includes object segmentation in images from a total of 80 objects categories .

To evaluate our proposal we intend to take as a baseline the training and retrieval done by only using repositories of real images. Since we intend to measure the advantages of including simpler images in the source domain, the evaluation would focus on the improvements achieved by using the abstract images dataset as a source domain and exploring domain adaptation on the task of image retrieval on real images. All tests are to be done using the same algorithm.

Describe the data you intend to use. Mention if there is an existing data source you intend to use, or if annotation is required

How will you evaluate your algorithm? What is a reasonable baseline?

Submission Instructions:

delete this section when submitting

You are required to use LaTeX to type your solutions to questions, and report of your programing as well. Other formats of submission will **not** be accepted. A template named "homework.tex" is also provided for your convenience.

After logging into data.cs.purdue.edu (physically go to the lab or use ssh remotely, you are all granted the accounts to CS data machines during this class), please follow these steps to submit your assignment:

1. Make a directory named *'your Name_your Surname'* and copy all of your files there.

2. While in the upper level directory (if the files are in /homes/dan/dan_goldwasser, go to/homes/dan), execute the following command:

   `turnin -c cs578 -p PROPOSAL *your_folder_name*`

   (e.g. your prof would use: `turnin -c cs578 -p PROPOSAL dan_goldwasser` to submit his work)

   Keep in mind that old submissions are overwritten with new ones whenever you execute this command.

3. You can verify the contents of your submission by executing the following command:

   `turnin -v -c cs578 -p PROPOSAL`

   Do **not** forget the -v flag here, as otherwise your submission would be replaced with an empty one.