UNIVERSITY OF JEAN MONNET

Laboratoire Hubert Curien

Connected Intelligence team

Object extraction techniques and visual image search with Semantic web techniques

Submitted by:
Aninda Maulik,
CPS2

Supervisor:
Prof. Pierre Maret
Dennis Diefenbach

June 25, 2020

Abstract

This internship is about exploration of object detection and extraction techniques with a state of the art computer vision api. Thereafter, we design a semantic web model for the extracted data and finally implement a visual image search engine through Qanswer.

1. Introduction

Users' experience is an important factor for the success of a given application. Thus, the front-end of Qanswer which highly impacts the users' experience, is an important part for a image base query system. Qanswer, well handles the translation from a natural language question to correct SPARQL queries.SPARQL has emerged as the standard RDF query language. An RDF query language is able to retrieve and manipulate data stored in Resource Description Framework (RDF) format. RDF data model is based on the idea of making statements about web resources in expressions of the form subject–predicate–object, known as triples. The subject denotes the resource, and the predicate denotes traits or aspects of the resource, and expresses a relationship between the subject and the object. We generate the rdf data model from a csv file, in which each line includes information for a triplet and all its components. The csv file is generated by consolidating the information and details about the required images. We primarily get the information of the required images by running the state-of-the-art, real-time object detection system; YOLO(You Look Only Once).

2. Presentation of the research problem

There is no way to use the knowledge generated by computer vision techniques, to query image bases.The research community has made a lot of efforts to use the computer vision techniques for extracting knowledge from images. On the other side, not much attention has been paid to the implementation of methods for making this knowledge available.We hope to change this trend by presenting Semantic Web techniques for querying the knowledge made available by computer vision. My work focuses on bridging the two disciplines here.

3. State of the art

Let's try to understand how does a Google image search engine work. There are two ways in this. 1) Indexing the text surrounding any image and matching it with the given query. If query matches, the corresponding linked image is retrieved. 2) Going ahead, in addition to linking text surrounding an image to that image, we can link all visually similar images to that image with the same text. e.g. consider an image/photo Img1 on any site with it's surrounding text Txt1. And lets say there are some other images Img2,Img3,Img4 etc. which may or may not have text but their (visual) content matches with the contents of Img1. Now for given query, if Txt1 is a good match, the retrieved result can contain Img1 in addition to Img2, Img3, Img4, etc. This is just one factor in addition to many other like matching query with text, features used to represent an image, page-rank of page containing an image, relevance, indexed database size available with search engine, etc. Huge indexed database availability with Google is one of the reasons why Google can give you best search results. Hence, when we ask for pictures of bicycle, we get many photos. However, when we try to search for images of bicycles on the left part of the photo, we get all the bicycle images which may or may not contain a photo of a left hand sided bicycle. Thus, we can conclude that google doesn't index their image base, based on object position. Qanswer,on the other hand, converts the natural language into triples and use the best ranked SPARQL query to query structured data sources. Hence, when we try to query/search for images of bicycles, we don't get much results; because the query is being made over structured data sources only and the existence of structured data is limited. But, the results are reliable. Now, if we try to search for pictures of bicycle on the left of the image, over Qanswer, then we're also going to get accurate results. The details are explained in the upcoming section. We would just like to say that, as discussed before, it would all start from converting the natural language text into triples; the triples would be then matched with an RDF file embedded within Qanswer. Thereafter, SparQL queries would be generated to make queries over structured data sources based on the RDF file information. All our efforts goes to the creation of this one RDF file which makes the difference. This RDF file gives the ability to Qanswer, to do object position detection in an image.This ability makes Qanswer, the best question answering system.

4. Contribution/Proposal description and implementation

This internship is an extension of the research topic 'Knowledge Elicitation via the Sequential Probabilistic Inference for High-Dimension Prediction' [1]. The existing model takes input from the expert who either has knowledge of the coefficients of the covariates or the relevance of the coefficients in the form of not-relevant, relevant or uncertain features. This input is incorporated in the feedback model and is used to sequentially update the posterior distribution. The model proposed here considers the pool of experts to be voters in a majority vote on each feature of the dataset.This majority voted knowledge is then incorporated in the form of feedback and is used to sequentially update the posterior distribution as before.

Algorithm 1 Knowledge Elicitation

```
Input: $D = (y_i, x_i) : i = 1, .., n$
Output: $p(f_{t+1}|D, f_1, .., f_t)$
Calculate $p(\theta|D)$ using
loop $(f = 1 : t)$
    Take feedback from the experts and do a majority vote on each feature for its inclusion or exclusion
    Sequentially update the posterior distribution to obtain $p(f_{t+1}|D, f_1, .., f_t)$
end loop
```

5. Experiments

The proposed model was tested on a dataset with Amazon products in the kitchen appliances category.

5.1. Dataset used

Amazon Data The Amazon data is a subset of the sentiment dataset of [2]. The dataset contains textual reviews and corresponding ratings 1-5 for Amazon products. Here, we have only considered reviews for products in the kitchen appliances category amounting to 5149 reviews. Each review is represented as a vector of features, for each distinct feature and each review, a matrix of occurrences was created and only those features were kept which appeared in atleast 100 reviews, a total of 824 features. For user study, ten university students and researchers were asked feedback on all the 824 features in the form of not-relevant, relevant or uncertain. It was assumed that the algorithm had access to 100 training data and at each iteration it could query the pre-given feedback one word at a time. The dataset was partitioned into three parts:

1. a training set of 100 randomly selected reviews
2. a test set of 1000 randomly selected reviews
3. a ”user-data-set” for constructing simulated user-knowledge [1].

5.1.1. Preprocessing of the Amazon dataset

The same pre processing techniques that were used by [1] were used here. At first the review file was read, the bag of words of the full data were prepared and the response values were extracted. Then,keywords with less than 100 occurrences in the reviews were removed. Finally, cross validation on the data was done to learn the best parameters for Spike-and-Slab linear regression. After that, the best parameters were used to fit a model to the data. The output was used as some kind of a ground truth of data.

5.2. Experimental results

The following are the mean squared errors for:

1. 10 individual real experts
2. 5 majority voted experts
3. 10 majority voted experts

The model parameters were set to π=0.7,ψ^2=0.01, $\alpha_\sigma = 1$,$\beta_\sigma = 1$, $\rho = 0.3$ (Eq. 4)

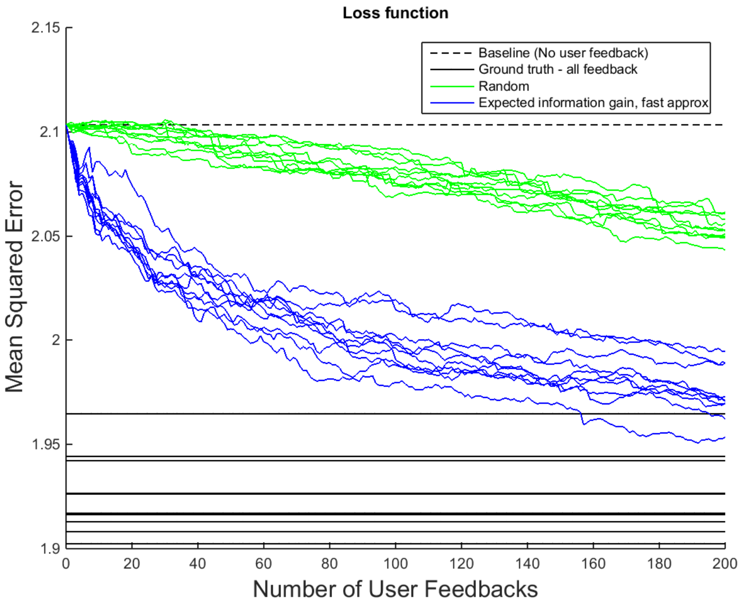

Figure 1. Mean squared error for 10 individual experts

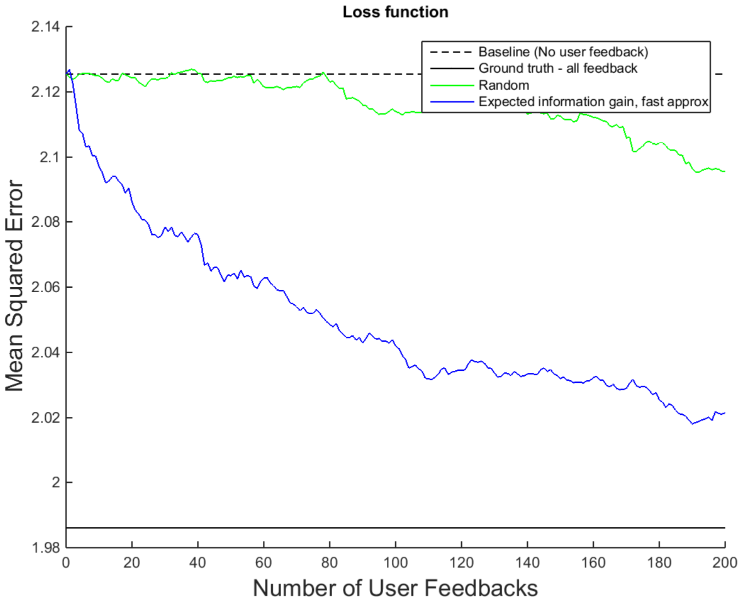

Figure 2. Mean squared error for 5 majority voted experts

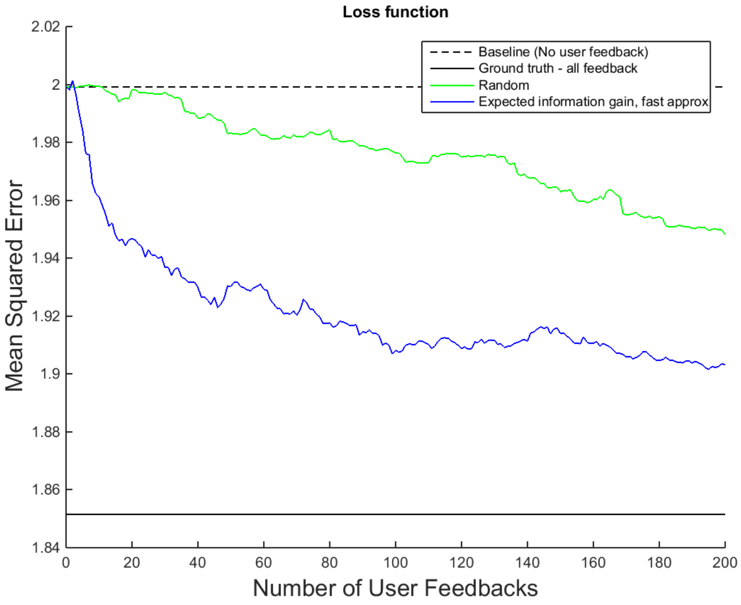

Figure 3. Mean squared error for 7 majority voted experts

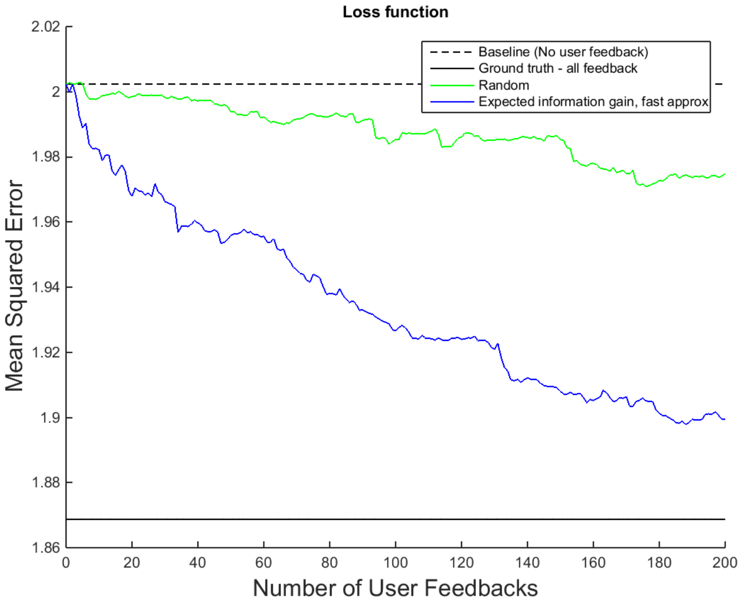

Figure 4. Mean squared error for 10 majority voted experts

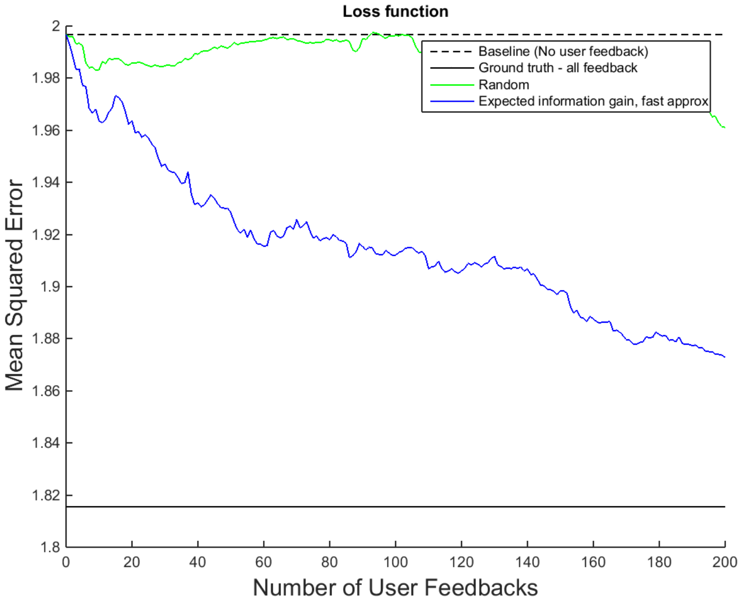

Figure 5. Mean squared error for 10 majority voted experts with π=0.8

It is clear that the use of expert knowledge via the sequential query model improves the average mean squared error at a better rate than random feature suggestion. The sequential model shows a faster rate of improvement than the random strategy which suggests that it asks about the most important features first.
While comparing the 5 majority voted, 7 majority voted and 10 majority experts, we see that for 5 majority voted experts the results were similar to individual experts being used for prediction, there was no real improvement in the mean square error values. However,

with increase in the number of majority voted experts to 7, there was a slight improvement in the MSE values and for 10 majority voted experts the improvement was quite visible.Also, there was also a sharper improvement in the MSE values with increase in the number of feedbacks from the very beginning for the 10 majority voted experts case. This seems to suggest that with increase in the number of experts, a majority vote strategy on the features seems to bring about better results. Also, with increase in the confidence coefficient of the experts there seems to be a slight improvement in the mean squared error.

6. Conclusion

A majority voted knowledge elicitation problem was presented for solving the ”small n, large p” problem in a sparse linear regression model. The model showed improved prediction accuracy(MSE reduction) even with just a few feedbacks when compared to the model with random feature selection. It also performed slightly better with increase in the number of experts for majority vote. Also, the prediction error reduced slightly when the confidence coefficient was increased. This method reduces the individual biases of the experts for better prediction accuracy. The presented method is generic and all the assumptions are probabilistic, thus it can be be tailored to suit other elicitation settings.

7. Future Works

A more extensive research can be done with more number of experts with different levels of expertise to study the behavior in the reduction of the mean squared error.

References

[1] Pedram Daee, Tomi Peltola, Marta Soare, and Samuel Kaski. Knowledge elicitation via sequential probabilistic inference for high-dimensional prediction. *CoRR*, abs/1612.03328, 2016.

[2] John Blitzer, Mark Dredze, and Fernando Pereira. Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification. In *In ACL*, pages 187–205, 2007.