Building Skin Condition Recogniser using Crowd-sourced High Level Knowledge

Orod Razeghi
http://www.cs.nott.ac.uk/~ozr
Hao Fu
hao.fu@nottingham.ac.uk
Guoping Qiu
qiu@cs.nott.ac.uk

VIPLAB, Computer Science
University of Nottingham
Nottingham, UK

Abstract

It is believed there are between 1000 to 2000 skin conditions of which 20% are difficult to diagnose. An intelligent diagnosing system not only helps patients with no or little access to health services, but also benefits typical general practitioners who have received minimal dermatology training. In this paper, we introduce a challenging dataset containing 2309 images from 44 different skin conditions. We employed 361 "Amazon Mechanical Turk" workers to answer some perceptual questions that represent the human understanding of these images. We present a novel random forest based "Human in the Loop" framework to efficiently fuse images' visual data and workers' answers for a better classification performance. We also suggest a new method to select the best sequence of questions to ask from the workers. Experiments demonstrate that this solution enhances classification accuracies, while minimising human unnecessary involvement.

1 Introduction

A recent comprehensive assessment of healthcare needs for skin conditions in the UK [6] suggests that 54% of the population experience a skin condition in a given twelve month period, and around 23% to 33% of the population have a skin problem that can benefit from medical care at any one time. The UK healthcare system relies on primary care as gatekeepers but typical general practitioners (GPs) paradoxically get minimal training in dermatology. Clearly, there is an acute skill shortage to meet the healthcare needs. A system that could automatically recognise at least life threatening skin conditions would be ideal. However, the state-of-the-art automatic computer techniques are still far from satisfying. A more realistic way is to utilise the human knowledge by including the human in the decision-making loop. This boosts accuracy of such system, and also helps with the issue of trust and public alienation towards autonomous technologies.

To realise this system, there are several core problems, which need to be tackled. Firstly, how to efficiently utilise these users provided information? Secondly, how to utilise these information in an online fashion? Thirdly, how to reduce the user workload? Finally, a relatively large scale dataset is necessary to evaluate the algorithm. In this work, we introduce a novel dataset[1] containing images and user provided information of various skin conditions.

[1] We have plans to release this dataset online in the future.

We also introduce a novel human in the loop framework based on random forests that efficiently fuses the two sources of information essential in solving this problem of fine-grained visual object classification. We emphasise human interactions with the system provides invaluable information that refine our recognition output but the burden on the user is kept to minimum by our ranking technique.

2 Related Work

There exists a fairly limited literature on "human in the loop" philosophy. The most similar work to ours may be [2], which propose to use a Bayesian framework to combine visual information and user provided answers for bird species recognition. However, it seems their Bayesian method struggles to fuse the two sources of visual and high-level human information, as each component in the framework is estimated separately and put together subsequently to form a recognition. This kind of late fusion does not consider the interactions between visual features and user answers. More importantly at each Question and Answer step, their Bayesian framework only considers a limited number of user answers, and there is no confident way to know when to stop asking new questions. In contrast, our proposed solution takes into account a full set of answers containing both user provided and automatically predicted answers. This allows the user to answer as much, or as little as she desires. Furthermore, their Bayesian solution could become computationally expensive. There are limited sensible assumptions to make it tractable, and this leads to its inflexibility.

Despite technological advancements, teledermatology (TD) and computer aided diagnosis (CAD) have had limited success. Most research in applying CAD to dermatology has been limited to melanoma conditions and using dermatoscopic images [7]. Surprisingly little research exists in recognition of ordinary photographical images. Wide availability of smart phone devices have spurt extensive activities to exploit these advancements. A dermatology-themed apps survey in [5] has come to conclusion that ubiquitous mobile computing offers new possibilities for help with patient care; however, all existing systems follow the traditional TD paradigm, and none have intelligent CAD capabilities.

One of few exceptions to the above is [8] that presents an interactive skin lesion recognition system based on a human in the loop visual recognition technology. In the paper, computer vision algorithms and models of human responses to a series of simple perceptual questions are combined together to achieve acceptable recognition rates. The proposed method utilises a similar Bayesian framework as in [2] with the same shortcomings, we discussed previously. They introduce a dermatology Q&A bank consisting of 21 questions and over 100 answers. However, their two "first" and "second" datasets contain only 3 and 7 skin conditions respectively, in contrast to our 44 classes. Moreover, their dataset includes only 796 images, in comparison to our 2309 skin condition images.

3 Implementation

3.1 Random Forest for Classification

Visual Representation: Image representation plays an important role in the quality of any classification solution. We have only utilised one feature in this work to represent visual information of each image but we believe that a combination of more features may improve accuracy of our algorithm. Our solution benefits from a visual feature that was proved to be very effective in similar datasets [9]. Pyramid Histogram of Visual Words (PHOW) [11] with specific parametrisation was extracted to form visual feature vectors of 1024 dimensions.

Nodes Split Function: Kernel PCA [10] is a suitable dimension reduction method to get a more compact representation for any chosen feature channel. We use kernel PCA to reduce our PHOW feature to a fixed low dimension.

User Answer Utilisation: We also utilise user provided information, which is in form of answers to perceptual questions, in our classification algorithm. These answers can be regarded as presence of tags[2] in each image. The importance of these answers become apparent when visual features fail to capture the complexity present in visually similar images. User provided answers can be used to build feature vectors with each element representing the presence of a tag. Instead of only 0 and 1 values, users' answers to the binary questions can be quantified by a certainty value, i.e. guessing, definitely, probably. These certainty values allow the framework to assign more weights to more confident answers. Each element in the vector is therefore set as a discrete probability between 0 to 1 representing the probability of a tag belonging to an image. Any positive answer has a probability value above 0.5, and any negative one is below 0.5. Table 3 shows these values.

Classification Method: Now we have defined methods to represent each image by a visual feature vector concatenated with its user answers vector. These answers vectors have a dimension of 37 representing the 37 questions in table 1. These concatenated vectors are used by a bootstrap aggregating (bagging) ensemble algorithm that trains 300 random trees. The information gain, calculated based on class labels of the training images, is used to select the best split function. Leaf nodes store a normalised probability distribution of the occurrence of all possible classes in the dataset. A common voting technique classifies the image.

3.2 Random Forest for Automatic Answer Prediction

The performance boost by the human in the loop is only valuable if the burden on the user is kept to the minimum. As the previous random forest is trained both on visual information and user provided answers, it becomes useless when the user answers only a subset of questions. We need to automatically predict responses for those unanswered questions. Unlike previous methods [2, 8], we treat this as an annotation problem where predicting presence of tags is the same as predicting answers. Not all automatic annotations will be perfect. Therefore, the least confidently predicted tags will be asked directly from the user. Sorting the prediction probability of tags in reverse order provides the algorithm with a ranking list of most important questions to ask from users.

[4] propose an interesting method that uses random forest for tag prediction. They use tag information instead of class label information to guide the generation of random trees. Thus, correlation among different tags is implicitly modelled. They also suggest two new concepts "Semantic Nearest Neighbour" and "Semantic Similarity Measure" that indicate "which" and "how many times" training images fall on the same leaf node with the query image. Based on their approach, we can automatically predict the existence of all possible tags or answer all questions. These predicted tags will be associated with a probability indicating how likely they are about to occur. More specifically, we denote I the query image and Q the probabilities of assigning tags. Let I_i represent I's ith semantic neighbour. Its count value is denoted as c_i. The ground truth tags of I_i is denoted as T_i. Suppose there are M tags in total,

[2] We will use tag(s) and answer(s) interchangeably in the rest of this paper.

Table 1: Dermatology Dataset Questions

Qs Group	Yes or No Binary Questions
Age	Infant, Child, Adult, Old
Site	Head, Mouth, Trunk, Arms, Sex Organs, Legs, Nails
Number	Single, Multiple
Distribution	Bilateral, Unilateral, Localised, ...
Arrangement	Discrete, Coalescing, Annular, ...
Type	Flat, Raised Solid, Fluid Filled, Broken Surface
Surface	Normal, Scale, Broken Surface, Changes in Thickness
Colour	Blood, Pigment, Lack of Blood, ...
Border	Well defined, Poorly defined
Shape	Round, Irregular

Table 2: Classification Accuracies

Feature	LIBSVM
Visual	13.37%
Tags	14.77%
Vis+Tags	16.03%
Feature	**Random Forest**
Visual	**15.46%**
Tags	**16.23%**
Vis+Tags	**21.69%**

Table 3: User Answers Certainties

Answer	Positive	Negative
Guessing	0.625	0.375
Probably	0.75	0.25
Definitely	1	0

hence Q and T_i can be represented as M size vectors: $Q = (q_1, ..., q_M)^T$ and $T_i = (t_{i1}, ..., t_{iM})^T$. Here t_{ij} is an indicator function that shows tag j probability for the ith image. The prediction of Q is totally influenced by the T_i and c_i value:

$$q_j = \sum_{i=1}^{K} \left(\frac{t_{ij}}{Z} \times f(c_i) \right), \; j \in \{1, 2, ..., M\} \tag{1}$$

Z is a normalizing constant, which is equal to $\sum_{i=1}^{K} \sum_{j=1}^{M} t_{ij}$. The term $f(c_i)$ represents a function that monotonically increases with c_i. $f(c_i)$ in our work is: $f(c_i) = c_i^2$.

4 Experiments

4.1 Dataset

We developed a challenging dataset over 3 months for this specific application. This dataset contains images of skin conditions from 44 different diseases. There are 880 training and 1429 testing images, totalling 2309 images. The lesions are manually segmented using a bounding box that includes pixels of lesion, healthy skin, and noise such as hair. Features are extracted from the entire bounding box, which as a whole is treated as a single instance. Images with their ground truth classification are from *http://www.dermis.net*. An Example of dataset image can be found in figure 2. Skin lesion images in our dataset range from different types of Eczema to various cancerous conditions, such as Superficial Spreading Melanoma.

The set of questions, which summarises the patient's skin lesion characteristics, are available in the dataset too. Medical professionals and a dermatological reference [1] were used to scientifically derive these questions and answers. The dataset contains 37 possible questions. Answers to simple perceptual questions were collected from 361 "Amazon Mechanical Turk" workers to form the database. Figure 2 represents a screenshot from the template used by the workers. Table 1 illustrates the type of questions and answers we used in our solution.

4.2 Results

Baseline Classification Accuracy: We employed LIBSVM (A Library for Support Vector Machines) [3] as a baseline to measure the quality of our random forest solution. The

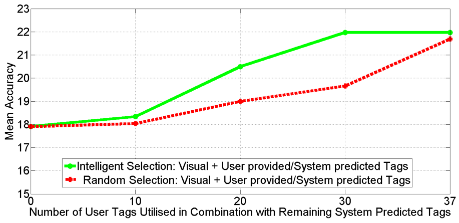
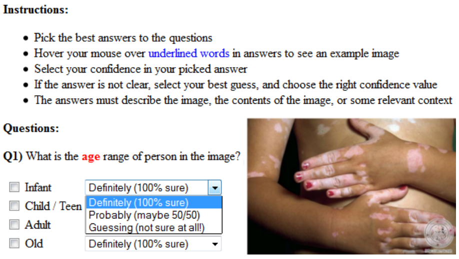


Figure 1: Mean classification accuracy results: System predicted tags reduce the number of user tags required to achieve peak performance. Results from randomly picked tags is also illustrated.



Instructions:

- Pick the best answers to the questions
- Hover your mouse over underlined words in answers to see an example image
- Select your confidence in your picked answer
- If the answer is not clear, select your best guess, and choose the right confidence value
- The answers must describe the image, the contents of the image, or some relevant context

Questions:

Q1) What is the age range of person in the image?

□ Infant — Definitely (100% sure)
□ Child / Teen
□ Adult
□ Old — Definitely (100% sure)

Figure 2: AMT interface used by workers. Image courtesy of: *http://www.dermis.net*

mean classification accuracy of LIBSVM over 5 runs using visual features, and tuned by default parameters levels at 13.37%. The LIBSVM classifier using tags features results in an accuracy of 14.77%. The combination of visual and tags features leads to a 16.03% mean accuracy. These baseline results illustrate the sheer difficulty of our dataset.

Random Forest Classification Accuracy: Our random forest trained by 300 trees and the same visual features results in an average accuracy of 15.46%. We also tried training the same number of trees only with tags features. The average accuracy saturates at 16.23%. Our random forest performs better than LIBSVM in both visual only and tags only cases. More importantly as it is clear, not the visual only nor the tags only results are accurate enough but once these features are combined, the classification accuracy rises to 21.69% using 300 trees. This shows the power of additional answers from users in samples where the visual features fail to capture the complexity of visually similar images. Table 2 summarises these results.

Automatic Answers Accuracy: It is very interesting to note that our solution is capable of answering all the questions automatically, and achieving a better performance than visual only results. Visual only features classification accuracy saturates at 15.46%, while the combination of these visual features with our fully predicted answers results in an average accuracy of 17.91%.

Questions Ranking Effect: It is imperative to clarify the fact that the user in our system doesn't need to answer all questions. Our model utilises both user provided answers, as well as automatically predicted tags in calculating the final results, despite the fact that some of these tags may have been wrongly predicted. Figure 1 represents the effect of adding user provided answers to our solution. As we gradually replace least confident automatic tags with user tags, the average accuracy rises. It is important to note that the system does not require to use all the user tags to achieve its peak performance. In the same figure, results from randomly picked tags is also presented. It is obvious that randomly picking user tags has not the same effective results as picking the least probable ones using our solution.

5 Conclusion

In this paper, we introduced a novel dermatology dataset. We proposed a random forest technique that combines heterogeneous data to achieve promising recognition rates. We also proposed an intelligent method to select the best sequence of questions that improves performance, while removing the burden on user's side.

References

[1] R. Ashton and B. Leppard. *Differential diagnosis in dermatology*. Radcliffe, 2005. ISBN 9781857756609. URL http://books.google.co.uk/books?id=TsbewCcanmgC.

[2] Steve Branson, Catherine Wah, Florian Schroff, Boris Babenko, Peter Welinder, Pietro Perona, and Serge Belongie. Visual recognition with humans in the loop. In *Proceedings of the 11th European conference on Computer vision: Part IV*, ECCV'10, pages 438–451, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-15560-X, 978-3-642-15560-4. URL http://portal.acm.org/citation.cfm?id=1888089.1888123.

[3] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[4] Hao Fu, Qian Zhang, and Guoping Qiu. Random forest for image annotation. In *12th European Conference on Computer Vision (ECCV)*, 2012. URL http://ima.ac.uk/papers/Hao2012c.pdf.

[5] A. Hamilton and R. Brady. Medical professional involvement in smartphone apps in dermatology. *British Journal of Dermatology*, (10.1111/j.1365-2133.2012.10844.x), Jan 2012.

[6] D. Grindlay J. Schofield and H. Williams. *Skin Conditions in the UK: a Health Care Needs Assessment*. Centre of Evidence-Based Dermatology, University of Nottingham, 2009.

[7] I. Maglogiannis and C. Doukas. Overview of advanced computer vision systems for skin lesions characterization. In *IEEE Trans Inf Technol Biomed*, pages 721–733, Sep 2009.

[8] Orod Razeghi, Guoping Qiu, Hywel Williams, and Kim Thomas. Skin lesion image recognition with computer vision and human in the loop. In *Medical Image Understanding and Analysis (MIUA), Swansea, UK*, pages 167–172, 2012. doi: 1-901725-45-6. URL http://ima.ac.uk/papers/Razeghi2012.pdf.

[9] Orod Razeghi, Guoping Qiu, Hywel Williams, and Kim Thomas. *Computer Aided Skin Lesion Diagnosis with Humans in the Loop*, pages 266–274. Machine Learning in Medical Imaging, 2012. doi: 10.1007/978-3-642-35428-1_33. URL http://ima.ac.uk/papers/Razeghi2012a.pdf.

[10] Bernhard Scholkopf, Alexander Smola, Klaus-Robert Muller, B Schölkopf, and KR Müller. Kernel principal component analysis. In *ICANN*, volume 1, January 1997.

[11] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.