Scene Categorization using Naive Bayes Classifier

Mahmut Demir

January 14, 2015

Abstract

In this work, we implemented a software that categorizes given image as one of predefined scene categories. Test user was presented a random image and then requested to specify some several objects that exist or do not exist inside a given image and our Naive Bayes based classifier categorizes image based on pre-constructed object-scene probabilities table. Conditional probabilities table created using open-source SUN image database [http://sundatabase.mit.edu/] which contains over 800 scene categories and more than 14.400 images and thousands of annotated images. For our case, we used 50 scene categories and 25 most common objects that exist in these images.

1 Introduction

Scene categorization is a fundamental problem in computer vision. Various approaches exist in the literature and because there is no general scheme in order to classify a scene into one category or some of the scenes can be classified into multiple categories it is wise to attack to the problem of scene categorization from the probabilistic viewpoint. Success rate of the classification algorithm highly depends on the amount and accuracy of the training data. Annotating training data is not just laborious task but also requires accuracy.

Open-source Scene UNderstanding (SUN) image database provides researchers with a comprehensive collection of annotated images covering a large variety of environmental scenes, places and the objects within. To build the core of the dataset, they counted all the entries that corresponded to names of scenes, places and environments using WordNet English dictionary. Once they established a vocabulary for scenes, they collected images belonging to each scene category using online image search engines by quering for each scene category term, and annotated the objects in the images manually.

A scene can be described in terms of a set of objects (e.g. car, window, tree, sky) that can be "generally" found inside and in terms of number of occurrences in the scene(one, several or many...). For example, we can describe "Street" scene as follows: Street scene must contain road,cars and sky on upper half along with sidewalks and buildings on both sides. There might be people, trees or road signs and so on... After enough training with representative images from various scene categories, we can create scene-object joint probabilities table. In such table, particular set of objects will be expected to have a higher probability of existence and some objects will have very low probability of presence given a particular scene category. In our work, we created such table using above-mentioned SUN image database and obtained probabilities of each object conditioned on scene category.

We used Naive Bayes Classifier (NBC) in order to classify given images based on objects detected within. Naive Bayes assumes a strong independence between objects given a particular scene category. Using this assumption problem is simplified greatly and inference is carried out using maximum a posteriori decision rule.

In order to test our approach, test user is provided with a random image and asked to choose a number objects that exist and do not exist in current image. Depending on evidence prompted from user, program outputs five best-matching scene categories.

2 Methods

Naive Bayes classifier is a simple probabilistic classifier based on Bayes' theorem with strong independence assumption between features given class label. It's proven that although Naive Bayes has simple probabilistic model, it has a very good performance with appropriate pre-processing and competitive results compared to other advanced methods such as Support Vector Machines. Naive Bayes has been preferred in many applications due to its simplistic and highly scalable nature and used extensively in document classification and medical diagnosis applications.

In our case, we have chosen Naive Bayes Classifier because it's fast, probabilistic model training is simple and it gives satisfying results even if test data is limited. Training the model requires to extract conditional probabilities of features conditioned on class labels. Because, tabular data for scene category-object occurrences does not exist, we first created a table in which most probable 50 scene categories and most common 25 objects included using information provided on SUN image database. Later, we calculated conditional probabilities of each object. Finally, in order to test our approach, we used an android based mobile application.

2.1 Extraction of statistics from SUN Image Database

SUN image database contains more than 130.000 images which are categorized based on 908 scene categories. There are more than 300.000 segmented and annotated objects with over 4000 object categories. Although SUN database offers tools to extract statistics from the database, we though it would be better to start with manageable number of scene and object categories. Therefore, we selected most common 50 scene categories and 25 common objects that can be generally found within typical scenes. We created a scene category-object occurrences table as shown in table 1 and class conditional probability calculations were done based on the created table.

There are several points to note when using the SUN database. First, note that each scene categories may contain different number of annotated images. Number of object occurrences depends on how many images annotated given scene category and they needs to be normalized before creating class conditional probabilities. Second, some scene categories do not contain enough annotated images therefore we would expect recognition accuracy to be low in these categories. Third, some objects have multiple occurrences and this leads the number of object occurrences to be higher than total number of annotated images for that category. This situation needs to be handled carefully in the calculation of class conditional probabilities.

2.2 Probabilistic Model

If we denote C_k as scene category k and e_i as evidence, problem then can be defined as follows:

$$p(C_k|e_1, e_2, ..., e_n) \;\; e_i \in \{o_j, \tilde{o_j}\} \tag{1}$$

$$where\ i = 1...N_{evidence},\ \ j = 1...N_{object},\ \ k = 1...N_{scene}$$

$p(e_i = o_j|C_k) \Rightarrow Probability\ of\ j^{th}\ object\ \underline{exist}\ in\ scene\ category\ k$
$p(e_i = \tilde{o_j}|C_k) \Rightarrow Probability\ of\ j^{th}\ object\ \underline{do\ not\ exist}\ in\ scene\ category\ k$

Problem here is that as total number of features, objects in our case, representing joint probabilities as a table becomes is infeasible and intractable. Using Bayes' theorem, above equation 1 can be equally written as:

$$p(C_k|e1, e_2, ..., e_n) = \frac{p(e_1, e_2, ..., e_n|C_k p(C_k)}{p(e_1, e_2, ..., e_n)} \tag{2}$$

We only interested in the numerator part of above equation, because denominator is just a constant and normalizer. Numerator is equivalent to joint probability:

$$p(C_k, e_1, e_2, ..., e_n) \tag{3}$$

Which can be written as follows, using chain rule:

$$\begin{aligned} p(C_k, e_1, e_2, ..., e_n) &= p(C_k)p(e_1, e_2, ..., e_n|C_k) \\ &= p(C_k)p(e_1|C_k)p(e_2, ..., e_n|C_k, e_1) \\ &= p(C_k)p(e_1|C_k)p(e_2|C_k, e_1)p(e_3, ..., e_n|C_k, e_1, e_2) \\ &= p(C_k)p(e_1|C_k)p(e_2|C_k, e_1)...p(e_n|C_k, e_1, ..., e_{n-1}) \end{aligned} \tag{4}$$

Conditional independence between features $e_i's$ given class label C_k assumption enables us to write above conditional probability as:

$$p(e_i|C_k, e_x, e_y, e_z) = p(e_i|C_k)\ for\ i \neq x, y, z \tag{5}$$

$$p(C_k, e_1, e_2, ..., e_n) = p(C_k)p(e_1|C_k)p(e_2|C_k)...p(e_n|C_k) \tag{6}$$

$$\begin{aligned} p(C_k|e_1, e_2, ..., e_n) &\propto p(C_k, e_1, e_2, ..., e_n) \\ &\propto p(C_k)p(e_1|C_k)p(e_2|C_k)...p(e_n|C_k) \\ &\propto p(C_k)\prod_i^n p(e_i|C_k) \end{aligned} \tag{7}$$

$$p(C_k|e_1, e_2, ..., e_n) = \frac{1}{Z}p(C_k)\prod_i^n p(e_i|C_k) \tag{8}$$

Hence, if we can obtain class conditional probabilities for each object, then we can calculate posterior probability of each scene category given some evidence about scene.

Formula we used in calculation of conditional probabilities can be basically stated as:

$$p(e_i = o_j|C_k) = \frac{N_{object_j k}}{N_{scene_k}} + \epsilon$$
$$= \frac{total\ \#\ of\ object\ type\ j\ present\ in\ images\ of\ scene\ category\ k}{total\ \#\ of\ images\ of\ scene\ category\ k} \tag{9}$$

One important point to note about above formula is that, if some objects during the training stage has never appeared for some categories, above conditional probability for that object becomes zero and it can cause to wipe out all information which comes from other probabilities. In order to prevent this, we introduced small correction term ϵ which ensures any probability will not be exactly zero.

Another important point to note is that,when some objects have multiple occurrences across some scene categories, above probability calculation may result in probabilities greater than 1.0 which is not possible. Quick remedy for that situation is to clamp probabilities between 0 and 1. However this might cause to loss some information and should be considered better solutions in order to recover all information.

2.3 Inference algorithm

After obtaining class conditional probabilities using formula 9, we calculated posterior probabilities using eq. 8 and normalized them over all scene categories. Decision rule we used for classification is maximum a posteriori approximation for finding most probable scene category given a number of evidences.

$$C_{\hat{k}} = \underset{k \in \{1 \dots K\}}{\arg\max}\, p(C_k) \prod_{i}^{N_{evidence}} p(e_i|C_k) \tag{10}$$

2.4 Software

Android mobile application was used for testing our algorithm. For this work, data table of 50 scenes and 25 objects shown in table 1 embedded in the software and conditional probabilities were inferred from the table using formula 1. Sample images which represent various scene categories obtained from SUN image database and embedded in software.

For the test scenario, user was prompted a scene image which was randomly selected from 20 sample images. First, user was requested to choose a number of objects out of 25 which exist in the current image. Next, user selected a number of objects which do not exist in the image. Five best matching scene categories with associated probabilities were calculated and shown on the screen after each evidence entry.

As we stated, scene category-object occurrences table embedded in the software and it is not possible for user to enter new training data or change present statistics. With future modifications in the software we can enable entering new training data which in turn increases success rate of classification. Moreover, only 25 objects were presented to user for evidence entry. Increasing number of objects recognized by the software will definitely increase success rate.

Scene	1 object	2 objects	3 objects	1 object	2 objects	3 objects
Bazaar Outdoor	Y(7.0)	N	N	N	N	N*(9.0)
Museum Outdoor	N(7.0)	N*(10.7)	Y(14.0)	Y(19.0)	Y(22.3)	N*(28.8)
Construction Site	N	N	N	N	N	N
Botanical Garden	Y(5.0)	Y(9.5)	Y(10.76)	Y(11.0)	Y(15.0)	Y(17.0)
Apartment Building	N	N	N	N	N	N
Building Facade	N	Y(6.8)	Y(7.0)	N	N	N
City	N*(5.0)	N*(6.5)	N*(13.39)	Y(21.0)	Y(39.9)	Y(41.0)
Downtown	N*(5.0)	N*(7.0)	N*(14.0)	Y(19.0)	Y(20.0)	Y(22.8)
Gas Station	Y(5.0)	Y(7.0)	Y(8.4)	N*(10.7)	N	N*(10.9)
Motel	N	N	N*(7.8)	N*(10.2)	N*(10.4)	N*(11.0)
Street	N	N	Y(16.0)	Y(36.0)	Y(44.0)	Y(62.0)
Schoolhouse	N	N*(7.0)	Y(9.6)	Y(11.2)	Y(20.6)	Y(20.7)
Convenience Store	Y(5.0)	Y(7.0)	Y(13.0)	Y(20.0)	Y(24.0)	Y(27.0)
Swimming Pool	N	Y(9.0)	Y(15.0)	Y(16.0)	Y(23.0)	Y(30.0)
Total Avg.	4/2.78	5/5.03	8/9.21	8/12.43	8/15.65	7/20.01

Table 1: Scene category recognition results. Y(x) = Category recognized successfully with probability x. N*(x) = Category not recognized but among five best matching. Last row shows recognition ratio along with probabilities of success.

3 Results

Results for our approach obtained using statistics from mobile application. Test user were presented 20 random images from the database and were allowed to enter 6 evidences, half of them were objects exist in the image and half were objects that do not exist. Results are summarized in table 1.

It was clearly observed that each additional evidence entry ”generally” increases success rate in terms of increasing probability of true scene category in the estimation. There are several important points which will be discussed further in the next section that affect the success rate of estimation:

1) Training data is not always enough! Sample image rarely contains objects not representative of scene category image belongs to.
2) Some objects(e.g. window, tree) have multiple occurrences in some scenes.
3) Some scene categories may contain similar objects with similar conditional probabilities.

4 Conclusion

As we stated in previous section, success rate of our algorithm was directly related to training data. In some of case, user is shown an image of some category but some objects within image might not be in the training data of that category. Second, some scene categories have identical object occurrence distribution with other categories which causes category of given image not recognized correctly. Third, some objects such as windows or trees in an image has multiple occurrences in images. Because these occurrences increase conditional probabilities, stating object existence or absence might not be enough. Instead, it would be wise to state number of occurrence of object either qualitatively(small, many) or quantitatively(1,2,3..).

In this work, we applied Naive Bayes Classifier to classify a given image into one of predefined scene categories. We used an open source image database which contains large amount of annotated images with over 900 scene categories. For our case, we applied categorization for 50 scenes. Classification was based on objects detected within image and class conditional probabilities were extracted from scene-object relation table we created. In order to test our algorithm, we developed an android mobile application in which test user was shown a random image and requested to enter some objects exists and do not exist within given image. After each evidence entry, best matching scene categories shown to user. Although results seem promising, success rate can be increase further by introducing more training data and more object categories.

References

[1] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba. *SUN Database: Large-scale Scene Recognition from Abbey to Zoo*. IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[2] Vasudevan, Shrihari, et al. Cognitive maps for mobile robots—an object based approach. *Robotics and Autonomous Systems* 55.5 (2007): 359-371.

[3] Naive Bayes Classifier, `http://en.wikipedia.org/wiki/Naive_Bayes_classifier`

		# of images	Wall	Windo w	Chair	Floor	Sky	Ceiling Lamp	Person	Buildin g	Trees	Ceiling	Tree	Car	Door	Cabinet	Plant	Table	Road	Mounta in	House	Ground	Grass	Car	Steps	SideW	Fence	Balcony
Outdoor Natural	Picnic Area	1		1			1			1	1		2					2					1					
	Playground	9		4			2		5	4	2		5		1				1				4					
	Beach	118					111		15	10	21		11							42								
	Forest	127		11			76		1		226		223				9			15		11						
Transportation	Backstairs	10		8	2		4			8	5		2		3		13					6						
	Bicyle Roads	11		6					11	6	3		3				5		3			4	5					
	Bridge	24					18			13	28		13	11			4		4	4		14	9					
	Bus Station	7		22			6		11	13	3				4				7				3					
	Gas Station	20	7	25			20	14	17	23	17		19	28	8				23				9					
Cultural and Historical	Bazaar Outdoor	10					6		46	13																		
	Campus	22	9	218					25	34	16		63		5		27		9					9	5			
	Castle	39	22	208			37		15	23	37		59		15		43			14	9	11	39					
	Fountain	15		39			7		13	15	6		12				11						13		4			
	Mosque Outdoor	17		15			16		86		13		53	9	13		44			7			5		7	10		
	Museum Outdoor	10	4	116			9		15	21			20	4	4		12		6				6		3	11		
Sports Field	Stadium Outdoor	6	3				6		2	7			6	8					4				6		2	3	2	
	Swimming Pool	23	15	34	16	23	19		17		25		15		8		19	8		8			12					
	Tennis Court	11		2			8			4	9		8		1							6	7				14	
Industrial	Construction Site	20	7	44			15		11	30			12	7					4							4		
	Factory Outdoor	10	5	51			10		3	15	11		9	9			7		7	2		6						
	Junkyard	10	3				8		3	2	11		4									6						
	Manhole	11				1				1									9		1	1	4		2	5		
	Rubble	10	3	2			9		7	16	4		4	1													2	
	Balcony exterior	18	8	22		1	11		51		23		2	6														19
	Botanical Garden	16					17		3		55		48				63			2	4	3	18					
	Garage Outdoor	14	5	14		2		2			8		23	2	3		12		6		18	9	7				2	
	House	18		109			12			4	5		18	4	15		28		11	9	28		20		10	11		
	Patio	7		10	21		2			3			5		4		20	4			4		2					
	Roof	11	1	3			7		4	8	6		2				6			1	1							
Commercial Buildings	Apartment Building	77	22	893			71		142	20			105	72	50		37		45							43		147
	Balcony outdoor	18	8	22		1	11	51		23		2	6	1	2													19
	Bazaar Outdoor	10							46																			
	Bistro Outdoor	6		23	19		4		8	12					4		4	17	7							7		15
	Building Comples	10	3	49			10		210	12	8		17	3			6		4			1	4			2	2	
	Building Facade	252		1468			157			396			112	223	323		49		158						45	143		253
	City	7		2			7			100	9		6	9					6				1			2		
	Convenience Store	12		14			11		7	21	3		4	17	13		4		12				3			8		
	Dinner Outdoor	14		81	4		11	12	8	22	12			22	11		13		12				3			13		
	Downtowm	11		52			10			83	23		14	11					12				3				3	
	Gas Station	20	7	25			20	14	17	23	17		15	28	8				23				9				8	
	General Store Outd	11	0	46			12		19		11		8	6	7				5			3	6		5	4		
	Hospital	14	5	75			14		6	19	8		18	16	2		14		8				12		5	12		
	Hotel Outdoor	12		101			8		19		4		15	11	3		7		5							10		5
	Motel	12		25			12		6	18	6		18	15	14		14		14	3	9					10		7
	Phone Booth	1		2						1																		
	School Hosue	8		39			8		1	10	12			1	9		6		1				10		2		2	1
	Shopping Mall Outd	1	0				1			1				3					1									
	Street	296	0	473					450	1405	85		187	1404					315	54						355		71

Figure 1: Scene Categories-Object Occurrences table

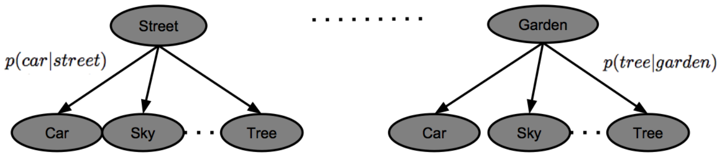

Figure 2: Graphical model of the problem. Note that objects are independent given scene category

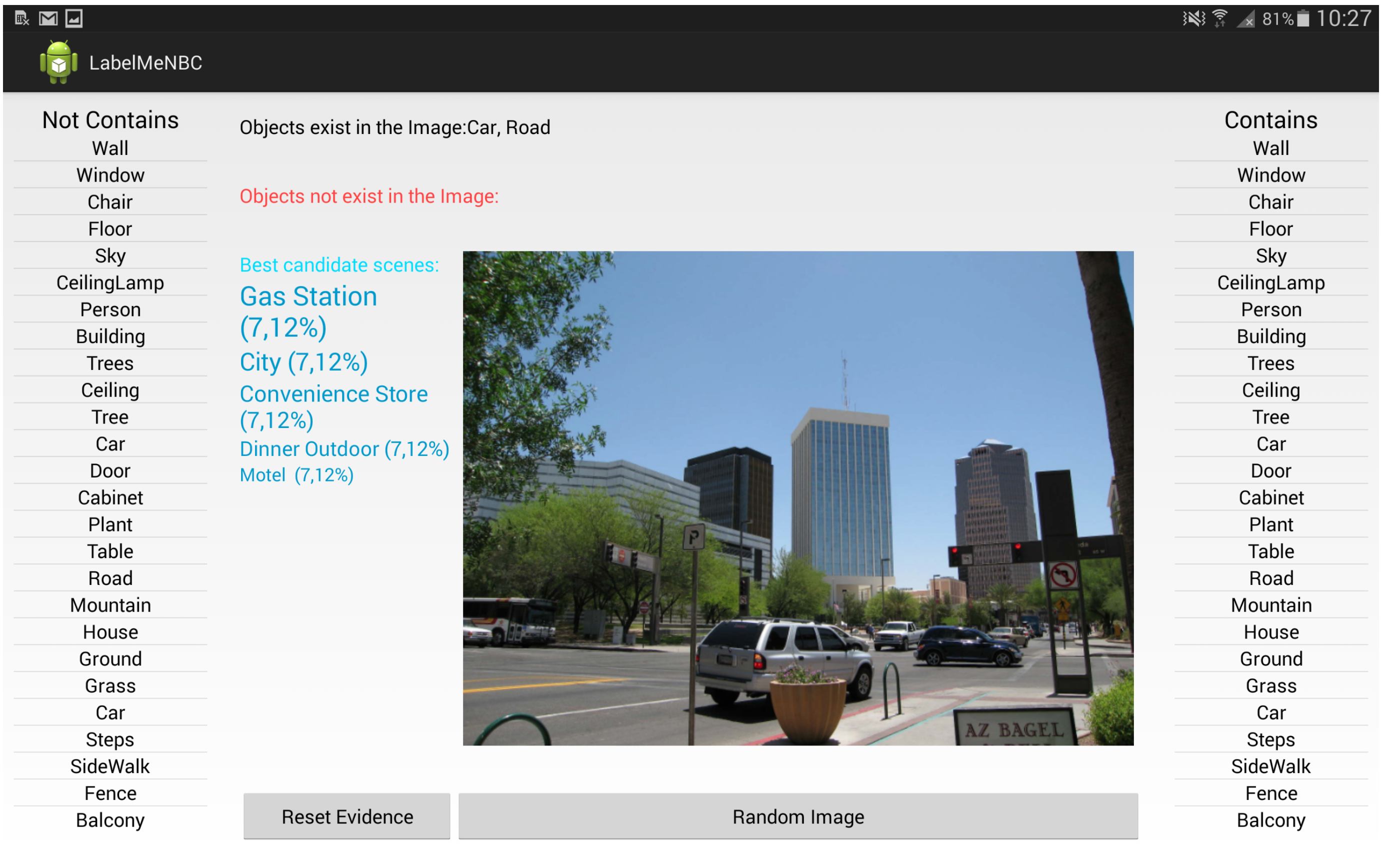

Figure 3: Mobile application test screen