

# Predicting Professions through Probabilistic Model under Social Context

**Ming Shao** and **Liangyue Li** and **Yun Fu \***

Northeastern University  
 Boston, MA, 02115

*mingshao@ccs.neu.edu, {liangyue, yunfu}@ece.neu.edu*

## Abstract

In this paper, we investigate the problem of predicting people's professions under social context. Previous work considering clothing information as well as fore/background context preliminarily proves the feasibility of predicting professions. In this paper, we discuss this problem in a more general case — multiple people in one photo with arbitrary poses, and argue that with appropriately built partial body features, spatial relations, and background context, more appealing results are achieved by a probabilistic model. We conduct experiments on 14 representative professions with over 7000 images, and demonstrate the model's superiority with impressive results.

## Introduction

In modern society, social status, connections, and people's roles in a particular situation draw great attention since they are fundamental elements of daily life. To automatically recognize the social roles, researchers propose to determine single person's demographical information by face at first, e.g., identity (Zhao et al. 2003), gender (Bourdev, Maji, and Malik 2011), age (Fu, Guo, and Huang 2010). Then more complex models considering pair-wise connections between people or context are introduced in (Naaman et al. 2005; Gallagher and Chen 2009; Wang et al. 2010; Berg et al. 2004; Xia et al. 2012).

In this paper, we describe an application scenario that can potentially boost the performance of social characteristics analysis — parsing the professions of people in a photo. People tend to make friends with those of the same professions, and any social website could utilize this for friend recommendation or professional services. We argue that the professions in a photo can be more precisely parsed by social context in a probabilistic model.

**Our contributions** First, we use poselet (Bourdev and Malik 2009; Bourdev et al. 2010), to capture the low-level feature, so we can deal with non-frontal upper body. Second, we use visual attributes (Farhadi et al. 2009; Kumar et al.

\*This research is supported in part by the NSF CNS award 1135660 and 1314484, Office of Naval Research award N00014-12-1-0125 and N00014-12-1-1028, Air Force Office of Scientific Research award FA9550-12-1-0201, and IC Postdoctoral Research Fellowship award 2011-11071400006.

Copyright © 2013, Association for the Advancement of Artificial Intelligence ([www.aaai.org](http://www.aaai.org)). All rights reserved.

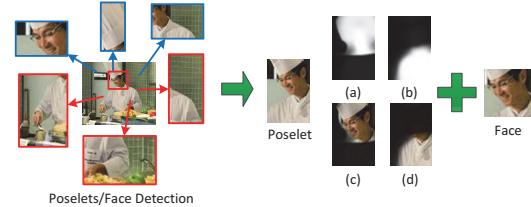


Figure 1: Poselets detection and masking. In the left figure, we detect several poselets and face from the center image. The poselets in blue frames are what we defined as “head poselets”, while those in red are “upper-body poselets”. In the right figure, we use different masks to post-process one poselet to remove irrelevant factors. (a) hat mask, (b) upper-body mask, (c) and (d) results after masking.

Table 1: Selections of poselet/region, mask type and feature type for each attribute. Note that “U-Body” means upper-body, and “Dense Grid” means the dense grid descriptions of HOG, LBP, and CIELAB color histogram. The numbers after “Uniform” and “Hat” denote the numbers of detailed attributes in these categories.

Attribute Name	Poselet/Region	Mask Type	Feature Type
Uniform (11)	U-Body Poselet	U-Body Mask	Dense Grid
Hat (7)	U-Body/Head Poselet	Hat Mask	Dense Grid
Skin Color	Body Region	N/A	Skin Feature
Skin Fraction	Body Region	N/A	Skin Feature
Age	Face Region	Face Mask	Raw Feature
Gender	Face Region	Face Mask	Raw Feature

2009), instead of low-level features as the input of person-level professions classifier. Third, we consider multiple people in a single photo and jointly determine their professions at the same time via a Bayesian network. This is inspired by the observation that people of correlated professions appear in the same photo with high probability.

## Data Representation via Attributes

Poselet body-part detectors (Bourdev and Malik 2009; Bourdev et al. 2010) and their relevant works have been successfully applied to many problems (Brox et al. 2011; Maji, Bourdev, and Malik 2011; Bourdev, Maji, and Malik 2011). In our framework, we mainly use poselets scheme for two tasks: **first**, to extract local body parts; **second**, to locate entire body region. We group the poselets into two sets,

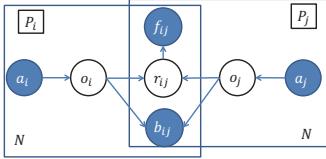


Figure 2: Graphical model of the profession prediction with social context.

Table 2: Notations of the graphical model.

$a_i$ : $i$ -th person attributes	$A$ : all persons attributes
$o_i$ : $i$ -th person profession	$O$ : all professions
$r_{ij}$ : relation type between two professions	$R$ : all relation types
$f_{ij}$ : spatial relation features	$F$ : all spatial relation features
$b_{ij}$ : shared background features	$B$ : all background features
$m_i = j$ : $i$ -th profession is assigned to $j$ -th person	$M$ : variable for assignments

namely, head poselets and upper-body poselets, as shown in Figure 1. Head poselets are responsible for attributes such as “hat”, while upper-body poselets for attributes such as “hat” or “uniform”. In addition, to remove noise, background and irrelevance, we design two masks to fit head and upper-body poselets, as in Figure 1.

We design 6 attributes to represent the semantic-level features that can fully or partially determine the professions of people in a photo. All attributes are listed in Table 1. We categorize them into two classes: (1) strong attributes; (2) weak attributes. Strong attributes can be utilized to determine the professions directly, i.e., uniforms, hats. Weak attributes, i.e., skin color and fraction, gender, age, are relevant factors to some professions but not sufficient.

We use SVM (Burges 1998) to learn the first four attributes listed in Table 1 by feeding it with features constructed by HOG (Dalal and Triggs 2005), LBP (Ahonen, Hadid, and Pietikäinen 2004), color histogram and skin fraction. We use the probabilistic outputs of SVM (Chang and Lin 2011) as the values of attributes. For the last two attributes  $\phi_5(x)$  and  $\phi_6(x)$ , i.e., gender and age, we compute through the framework in (Gallagher and Chen 2009), and quantize them into bins to formulate visual attributes.

## Predicting Professions with Context

We explain how to utilize contexts in a photo to predict multiple people’s professions. We use three kinds of contexts, namely, (1) spatial relations, (2) co-occurrence, and (3) background information. Based on the analysis, our problem is finally formulated as the graphical model in Figure 2 and its relevant notations are shown in Table 2. Our aim is to maximize the following conditional probability  $p(O, R|A, F, B)$ , which can be written as:

$$\frac{p(O, F, A, R, B)}{p(A, F, B)} \sim \sum_M p(O, F, A, R, B|M)p(M). \quad (1)$$

According to the graphical model, the former expression can be decomposed into the following formulation:

$$\sum_M \prod_i p(o_i|a_{m_i}) \prod_{i,j} p(f_{m_i m_j}|r_{ij}) p(b_{ij}|o_i, o_j) p(r_{ij}|o_i, o_j) p(M),$$

which can be estimated by EM algorithm. After that, we use the following objective function to infer on a test sample:

Table 3: Average precision (%) for attributes of uniform.

Uniform	CF	CG	CL	DT	FF
AP	64.8	48.7	57.6	42.2	45.6
Uniform	LY	MM	PM	SD	WT
AP	65.1	29.9	42.9	59.2	32.8

Table 4: Average precision (%) for attributes of hat.

Hat	DT	SD	CF	PM	FF	CL	None
AP	47.5	65.8	95.4	67.3	46.0	31.5	66.0

Table 5: Average precision (%) for weak attributes (WA).

WA	SC	SF	GD	(0-18)	(18-40)	(40-60)
AP	78.1	65.4	66.9	71.3	63.5	42.1

Table 6: Experimental results of average precision (%). The **average performance** of these methods are: Background Features (15.4%), Method in (Song et al. 2011) (35.0%), Person-Level Prediction (36.6%), Ours (40.4%).

	CF	CG	CL	CT	DT
Background Features	10.3	10.8	11.4	7.4	9.6
Method in (Song et al. 2011)	40.8	34.2	42.8	19.2	44.9
Person-Level Prediction	40.6	34.6	43.7	21.3	45.2
Joint Prediction ( <b>Ours</b> )	42.3	35.1	46.1	27.6	48.9
	FF	LY	MM	MR	PM
Background Features	8.3	31.7	19.7	12.8	9.1
Method in (Song et al. 2011)	31.3	59.1	21.8	48.2	18.4
Person-Level Prediction	30.3	57.6	23.1	52.1	20.1
Joint Prediction ( <b>Ours</b> )	36.7	60.1	27.4	57.3	21.5
	SP	SD	ST	TC	WT
Background Features	28.8	31.5	14.8	7.8	17.6
Method in (Song et al. 2011)	48.2	60.1	21.5	13.6	20.6
Person-Level Prediction	57.1	68.9	20.1	13.0	21.5
Joint Prediction ( <b>Ours</b> )	60.2	74.7	25.0	15.2	28.6

$$M^* = \arg \max_M p(M|O, F, A, R, B). \quad (2)$$

## Experimental Results

we collect more than 7000 images of 14 different professions from the websites, e.g., Google Image. This database will be online **publicly available** soon. First, we evaluate the performance of the proposed attributes. For each attribute, we use half of images with/without this attribute as training and the other half as test. We finally obtain the average precisions for each attribute shown in Table<sup>1</sup> 3, 4, 5, including strong and weak attributes classification results.

Second, we evaluate the proposed probabilistic model in Table 6, where four methods are compared with each other. The person-level prediction is similar to the method in (Song et al. 2011), but uses attributes as input. For the background features, we use SIFT + BoWs model (Fei-Fei and Perona 2005) to train a multi-class SVM for all profession categories, and use background features in test images as inputs. We can see that the proposed framework works comparably with the state-of-the-art method and sometimes even better, especially when multiple people are in a photo.

<sup>1</sup>Note we use abbreviation for each profession: CF-chef, CG-clergy, CL-construction labor, CT-customer, DT-doctor, FF-fire fighter, LY-lawyer, MM-mailman, MR-marathoner, PM-policeman, SP-soccer player, SD-soldier, ST-student, TC-teacher, WT-waiter, SC-skin color, SF-skin fraction, GD-gender.

## References

- Ahonen, T.; Hadid, A.; and Pietikäinen, M. 2004. Face recognition with local binary patterns. In *European Conference on Computer Vision*, 469–481.
- Berg, T. L.; Berg, E. C.; Edwards, J.; Maire, M.; White, R.; whye Teh, Y.; Learned-miller, E.; and Forsyth, D. A. 2004. Names and faces in the news. In *IEEE Conference on Computer Vision and Pattern Recognition*, 848–854. IEEE.
- Bourdev, L., and Malik, J. 2009. Poselets: Body part detectors trained using 3d human pose annotations. In *International Conference on Computer Vision*.
- Bourdev, L.; Maji, S.; Brox, T.; and Malik, J. 2010. Detecting people using mutually consistent poselet activations. In *European Conference on Computer Vision*.
- Bourdev, L.; Maji, S.; and Malik, J. 2011. Describing people: Poselet-based attribute classification. In *International Conference on Computer Vision*.
- Brox, T.; Bourdev, L.; Maji, S.; and Malik, J. 2011. Object segmentation by alignment of poselet activations to image contours. In *IEEE International Conference on Computer Vision and Pattern Recognition*.
- Burges, C. 1998. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery* 2(2):121–167.
- Chang, C.-C., and Lin, C.-J. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2:27:1–27:27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Dalal, N., and Triggs, B. 2005. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, 886–893. IEEE.
- Farhadi, A.; Endres, I.; Hoiem, D.; and Forsyth, D. 2009. Describing objects by their attributes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1778–1785. IEEE.
- Fei-Fei, L., and Perona, P. 2005. A bayesian hierarchical model for learning natural scene categories. In *IEEE Conference on Computer Vision and Pattern Recognition*, 524–531.
- Fu, Y.; Guo, G.; and Huang, T. 2010. Age synthesis and estimation via faces: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(11):1955–1976.
- Gallagher, A. C., and Chen, T. 2009. Understanding images of groups of people. In *IEEE Conference on Computer Vision and Pattern Recognition*, 256–263. IEEE.
- Kumar, N.; Berg, A.; Belhumeur, P.; and Nayar, S. 2009. Attribute and simile classifiers for face verification. In *International Conference on Computer Vision*, 365–372. IEEE.
- Maji, S.; Bourdev, L.; and Malik, J. 2011. Action recognition from a distributed representation of pose and appearance. In *IEEE Conference on Computer Vision and Pattern Recognition*, 3177–3184. IEEE.
- Naaman, M.; Yeh, R.; Garcia-Molina, H.; and Paepcke, A. 2005. Leveraging context to resolve identity in photo albums. In *ACM/IEEE-CS joint conference on Digital libraries*, 178–187.
- Song, Z.; Wang, M.; Hua, X.; and Yan, S. 2011. Predicting occupation via human clothing and contexts. In *International Conference on Computer Vision*.
- Wang, G.; Gallagher, A.; Luo, J.; and Forsyth, D. 2010. Seeing people in social context: Recognizing people and social relationships. In *European Conference on Computer Vision*, 169–182. Springer.
- Xia, S.; Shao, M.; Luo, J.; and Fu, Y. 2012. Understanding kin relationships in a photo. *IEEE Transactions on Multimedia* 14(4):1046–1056.
- Zhao, W.; Chellappa, R.; Phillips, P.; and Rosenfeld, A. 2003. Face recognition: A literature survey. *Acm Computing Surveys (CSUR)* 35(4):399–458.