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 foreground/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. 2010), to extract local body parts; third, to extract local body parts; second, to locate entire body region. We group the poselets into two sets,
**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:

\[ p(O, R|A, F, B) \sim \sum_{M} p(O, R, A, F, B|M)p(M). \]  

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

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

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

\[ M^* = \arg \max_{M} p(M|O, F, A, R, B). \]  

**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 precision for each attribute shown in Table 6, 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.

**Prediction of Professionals with Social Context**

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:

\[ p(O, R|A, F, B) \sim \sum_{M} p(O, R, A, F, B|M)p(M). \]  

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

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

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

\[ M^* = \arg \max_{M} p(M|O, F, A, R, B). \]  

**Predicting Professions with Social 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:

\[ p(O, R|A, F, B) \sim \sum_{M} p(O, R, A, F, B|M)p(M). \]  

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

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

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

\[ M^* = \arg \max_{M} p(M|O, F, A, R, B). \]  

**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 precision for each attribute shown in Table 6, 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.
References


