International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Practical learning approaches for undirected graphical models. Application to scene object recognition

Jose-Raul Ruiz-Sarmiento *, Cipriano Galindo, Javier Gonzalez-Jimenez

Machine Perception and Intelligent Robotics Group, System Engineering and Automation Department, University of Málaga, Campus de Teatinos, 29071, Málaga, Spain

A R T I C L E I N F O

Article history:
Received 14 May 2016
Received in revised form 5 August 2016
Accepted 24 October 2016
Available online xxxx

Keywords:
Undirected graphical models
Conditional random fields
Parameters learning
Training
Scene object recognition

A B S T R A C T

Probabilistic Graphical Models (PGMs) in general, and Undirected Graphical Models (UGMs) in particular, become suitable frameworks to capture and conveniently model the uncertainty inherent in a variety of problems. When applied to real world applications, such as scene object recognition, they turn into a reliable and widespread resorted tool. The effectiveness of UGMs is tight to the particularities of the problem to be solved and, especially, to the chosen learning strategy. This paper presents a review of practical, widely resorted learning approaches for Conditional Random Fields (CRFs), the discriminate variant of UGMs, which is completed with a thorough comparison and experimental analysis in the field of scene object recognition. The chosen application for UGMs is of particular interest given its potential for enhancing the capabilities of cognitive agents. Two state-of-the-art datasets, NYUv2 and Cornell-RGBD, containing intensity and depth imagery from indoor scenes are used for training and testing CRFs. Results regarding success rate, computational burden, and scalability are analyzed, including the benefits of using parallelization techniques for gaining in efficiency.

1. Introduction

Intelligent cognitive agents, *e.g.* mobile robots, aiming to successfully operate in human environments require the ability to *understand* what is going on in their surroundings. Scene object recognition systems are cornerstone components for such ability, providing the robot with high-level information that can be used for tasks like *scene understanding* [1,2] or *semantic mapping* [3,4]. The limited resources normally available in these agents force recognition systems to perform efficiently, while also dealing with the uncertainty latent in both, the robots' sensory system and their models of the working environment.

Probabilistic Graphical Models (PGMs) [5] in general, and Undirected Graphical Models (UGMs) in particular, also known as Markov Random Fields (MRFs) [6], offer suitable frameworks to tackle such uncertainty, incorporating contextual relations among the scene objects. Briefly, they rely on a graph representation to model the perceived objects as random variables in the form of nodes, and the relations among them as edges (see Fig. 1). Over this graph model, object recognition can be efficiently conducted by means of *probabilistic inference* queries. Previous to such inference, a *learning phase* must be completed in order to tune the model numerical parameters, also called weights, for the application at hand.

* Corresponding author.
E-mail addresses: jotaraul@uma.es (J.-R. Ruiz-Sarmiento), cipriano@ctima.uma.es (C. Galindo), javiergonzalez@uma.es (J. Gonzalez-Jimenez).

http://dx.doi.org/10.1016/j.ijar.2016.10.009
0888-613X/

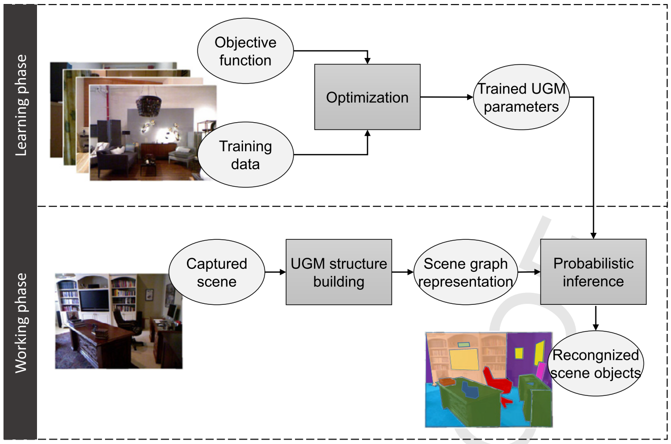

Fig. 1. Learning and working phases involving Undirected Graphical Models for scene object recognition. Boxes are processes, while ovals represent consumed/produced data. This work focuses on the selection of different objective functions and optimization techniques for scene object recognition using training data from the NYUv2 and Cornell-RGBD datasets.

Typically, the learning phase in MRFs is targeted at maximizing the *expected likelihood* of the model with respect to a set of training data. However, computing this likelihood requires *exact inference*, which is in general a $\mathcal{NP}$-hard problem [5,7]. Two major approaches stand out to overcome this concern: (i) the definition of alternative, tractable objective functions, or (ii) the estimation of the likelihood by approximate inference algorithms [8–10]. The performance of methods from both options highly differs depending on the domain of the problem at hand, *i.e.* the nature and internal structure of the data to work with. Therefore, for a given application, a thorough study is needed in order to obtain a successful model.

In this work we present a review of the most resorted learning approaches and empirically analyze their performance in the scope of the scene object recognition. The aim of this study is to serve as a guide to quickly set-up a working-system as successful as possible for such problem. Concretely, we focus on the analysis of the following objective functions for tuning the parameters of Conditional Random Fields (CRFs) [11,12], the discriminative variant of MRFs:

- The *pseudo-likelihood* function [13], as an alternative to the expected likelihood, and
- The most popular approximate inference algorithms for estimating the likelihood, including *Marginal* queries (Sum-product Loopy Belief Propagation [14]) and *Maximum a Posteriori* (MAP) queries (Iterated Conditional Modes [13], Graph-cuts [15], Max-product Loopy Belief Propagation [16]).

To complete the learning phase, an optimization process is needed to estimate the model parameters. Again, there is not an *ultimate* method, given that their performance depends on the chosen objective function to optimize and the problem domain, so two alternatives successfully applied in the related literature are explored in this work: the Stochastic Gradient Descent (SGD) [17] method, and the quasi-Newton Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) [18] one.

In order to test the trained CRFs, thorough evaluations are carried out using the Undirected Probabilistic Graphical Models in C++ library (UPGMpp) [19] and two widely-used datasets in robotics: NYUv2 (New York University Depth Dataset version 2) [20] and Cornell-RGBD (Cornell University RGB-D Dataset) [21]. Both datasets contain labeled intensity and depth imagery from indoor scenes, albeit they show distinctive characteristics whose influence in the learning phase is studied: while NYUv2 comprises a high number of labeled images (we have used 208 from home environments) that capture the objects and relations from portions of scenes (see Fig. 2-middle), Cornell-RGBD provides a lower number of scenes (28 from homes) but fully covering the inspected place (see Fig. 2-left), which results in a considerably larger number of perceived objects and relations.

The presented study focuses on two facets of the learning methods: the recognition performance of the trained CRFs, and the required computational time. To measure the CRFs performance we have executed different MAP inference methods over the learned models, and compared their recognition results with the ground-truth information provided by the datasets. The combination of Marginal inference and SGD for learning yielded the best recognition results, achieving a success of $\sim 80\%$ and $\sim 67\%$ in NYUv2 and Cornell-RGBD respectively. The computational time needed by each learning method to converge is also analyzed, studying the advantage of parallelization techniques. Results of the achieved speed-up are shown employing the Open Multi-Processing API (OpenMP) [22]. Finally, the scalability of the learning methods according to different factors is also studied.

2. Related work

In the last decade, the utilization of Probabilistic Graphical Models (PGMs) [5] for tackling the scene object recognition problem has become increasingly popular. This is, to a great extent, due to their suitability to efficiently model this type of