Conditional Random Fields for Modeling Structured Data

N

Tom Burr
Statistical Sciences, Los Alamos National Laboratory, USA

Alexei Skurikhin
Space Data Systems, Los Alamos National Laboratory, USA

INTRODUCTION

In statistical science or machine learning, pattern recognition is the task of using measurements from an input domain X to predict their labels from an output domain Y. A typical setting is binary classification where $Y=\{0,1\}$, with 0 being the label for "objects of type A" and 1 being the label for "objects of type B." For example, measurements of sepal and petal length and width from each of many iris plants have been used to predict the species from among the three species of iris studied by Fisher (1936). The inferred label for a specific sample of iris plants as one of three types is independent of the inferred label for another specific sample so this task does not have structure (defined below) in the output domain.

In general, there can be dependence among samples in the input domain, or output domain, or both or neither. For example, fraud detection in credit card transactions relies partly on pattern recognition methods that exploit relations among the samples (transactions), both in the input and output domains (Bolton & Hand, 2002). Pattern recognition is commonly applied to unstructured data for which the samples are independent with respect to both the input and output domains, as in Fisher's iris data. In contrast, when there is structure in the input and output domain, i.e., sample input and sample labels are not independent, the problem is referred as structured machine learning or structured prediction. Examples, among many others, include fraud detection in credit card transactions, labeling pixel patches in images and named entity recognition (NER) in text. NER is the task of identifying and classifying proper names in text, including locations (e.g., New York, China), names of people, companies and organizations. One challenge in NER is that correctly recognizing named entities requires context – information on co-occurrence of other entities in the surrounding text. Similar situations often occur in higher-level image interpretation. For example, recognizing a scene as an office environment increases the expectation of detecting a computer inside this scene; in contrast, the expectation of a computer inside a grocery store scene is not high as in an office scene.

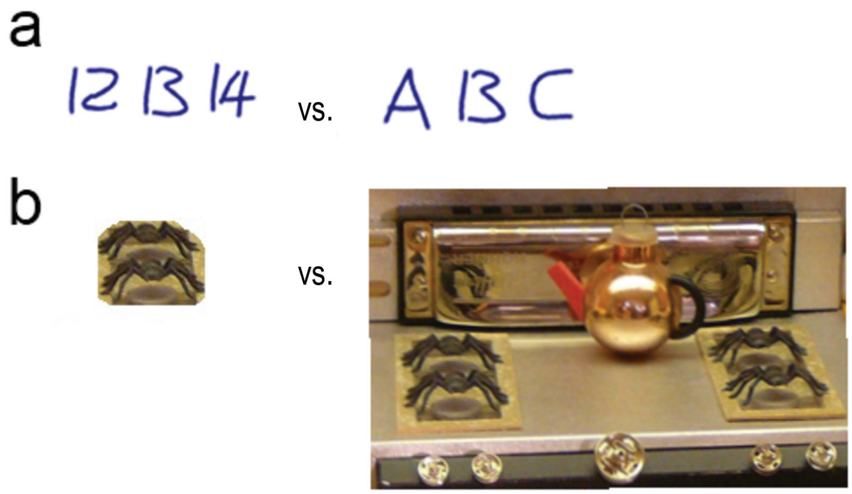
In addition to providing some general discussion, this article mainly uses image analysis applications in which there is dependence among samples (pixels or pixel patches) in the output domain because nearby pixels tend to have similar labels such as "natural" or "manmade," as explained below. Figure 1 illustrates the role of context in disambiguating object recognition. In Figure 1, a spatial context illustrates the central element, 13 in the context of numbers versus B in the context of letters (Bruner & Minturn, 1955) and in Figure 1b, what may appear as spiders, when viewed alone out of the surrounding context, becomes the top burner grates for gas range, when viewed in the context of the whole scene.

Probabilistic graphical models (PGMs) are being increasingly used to model problems having a structured domain. PGMs are represented by two main categories of models. Directed graphical models are known as Bayesian Networks (BNs) and undirected graphical models are known as Markov Random Fields (MRF) and Conditional Random Fields (CRF). PGMs are used to express dependencies between the input and output domains as well as dependencies within domains and to enable probabilistic inferences such as answering queries using the probabilistic model (e.g., a model based on a CRF or a MRF or a BN) of the problem (such as NER in text, image segmentation, or object recognition in images). A key task is to compute the

DOI: 10.4018/978-1-4666-5888-2.ch608

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Figure 1. An illustration of context to improve recognition of:(a) the central element, 13 in the context of numbers vs. B in the context of letters; (b) what may appear as spiders, when viewed alone out of the surrounding context, becomes the top burner grates for gas range, when viewed in the context of the whole scene



probability distribution over the variables of interest (for a test sample called the query) given the observed values of other random variables (the evidence).

CRFs were introduced by Lafferty et al. (2001) and have been successfully applied in areas such as video content analysis (Reiter et al., 2007), bioinformatics (Sato & Sakakibara, 2005), computer vision (Kumar & Hebert, 2003; He, 2004), and natural text processing (Lafferty et al., 2001; Sha & Pereira, 2003). While a MRF models P(*Y*,*X*) and includes a model for *P*(*x*), a CRF directly models a conditional distribution P(*Y*|*X*=*x*) without specifying a model for *P*(*x*). In general, modeling *P*(*x*) is difficult because feature functions between parts of the PGM model are often correlated. As a result of direct modeling of conditional distributions, CRFs have been found better suited to processes rich, dependent (overlapping) features without having to model *P*(*x*).

Inference in CRF models, which try to accurately represent real-life phenomena and corresponding distributions, is fundamentally intractable, unless the models have tree-like structures. Probabilistic inference in CRFs based on trees is efficient and exact. A challenge in CRF model structure construction and parameter optimization (learning) is to balance the expressive power of the models with the computational complexity of inference in the models, because the inference has to be performed both during model learning and for answering probabilistic queries based on the optimized model. This requires development of better approximate inference algorithms, and algorithms to learn the structure and parameters of the PGM model. These have the potential to significantly improve the quality of answers to probabilistic queries. Most effort has been devoted to approximate inference algorithms, which can be broadly categorized as Markov Chain Monte Carlo algorithms (MCMC) (Asuncion et al., 2010a, 2010b; He et al., 2004; Hinton, 2002; Carreira-Perpiñán & Hinton, 2005), that attempt to sample from the distribution of interest, and variational algorithms, that convert the inference problem into an optimization problem, which is then approximated until it becomes tractable (e.g., loopy belief propagation [Pearl, 1988; Frey & MacKay, 1997; Murphy et al., 1999; Yedidia et al., 2005] and graph cut [Boykov et al., 2001; Boykov & Kolmogorov, 2004; Kolmogorov & Zabih, 2004]). Much research has concentrated on the second category, because the first one is computationally more demanding.

Formally, if x_c is a feature vector, and y_c is the corresponding labels, a flexible model for the conditional probability is

$$P\left(y \mid x\right) = \frac{1}{\mathrm{Z(x)}} \prod_{c \in C} \psi_c(y_c, x_c, \theta),$$

and

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/conditional-random-fields-for-modeling-structured-data/113074

Related Content

Intelligent System of Internet of Things-Oriented BIM in Project Management
Jingjing Chen (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-14).*
www.irma-international.org/article/intelligent-system-of-internet-of-things-oriented-bim-in-project-management/323803

Meta-Context Ontology for Self-Adaptive Mobile Web Service Discovery in Smart Systems
Salisu Garba, Radziah Mohamadand Nor Azizah Saadon (2022). *International Journal of Information Technologies and Systems Approach (pp. 1-26).*
www.irma-international.org/article/meta-context-ontology-for-self-adaptive-mobile-web-service-discovery-in-smart-systems/307024

Facilitating Inclusive Teaching and Learning Spaces Through Digital Education Technology: Teaching and Learning Though Digital Technology
Tsediso Michael Michael Makoelleand Michelle Irene Somerton (2019). *Educational and Social Dimensions of Digital Transformation in Organizations (pp. 43-64).*
www.irma-international.org/chapter/facilitating-inclusive-teaching-and-learning-spaces-through-digital-education-technology/215135

BYOD (Bring Your Own Device), Mobile Technology Providers, and Its Impacts on Business/Education and Workplace/Learning Applications
Amber A. Smith-Ditizioand Alan D. Smith (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 5981-5991).*
www.irma-international.org/chapter/byod-bring-your-own-device-mobile-technology-providers-and-its-impacts-on-businesseducation-and-workplacelearning-applications/184299

Context-Aware Personalization for Mobile Services
Abayomi Moradeyo Otebolakuand Maria Teresa Andrade (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 6031-6042).*
www.irma-international.org/chapter/context-aware-personalization-for-mobile-services/184303