Available online at www.sciencedirect.com

ScienceDirect

Computer Speech & Language 45 (2016) 552–569

www.elsevier.com/locate/csl

Dialogue manager domain adaptation using Gaussian process reinforcement learning

Milica Gašić*, Nikola Mrkšić, Lina M. Rojas-Barahona, Pei-Hao Su, Stefan Ultes, David Vandyke, Tsung-Hsien Wen, Steve Young

University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK

Received 15 April 2016; Accepted 1 September 2016

Abstract

Spoken dialogue systems allow humans to interact with machines using natural speech. As such, they have many benefits. By using speech as the primary communication medium, a computer interface can facilitate swift, human-like acquisition of information. In recent years, speech interfaces have become ever more popular, as is evident from the rise of personal assistants such as Siri, Google Now, Cortana and Amazon Alexa. Recently, data-driven machine learning methods have been applied to dialogue modelling and the results achieved for limited-domain applications are comparable to or out-perform traditional approaches. Methods based on Gaussian processes are particularly effective as they enable good models to be estimated from limited training data. Furthermore, they provide an explicit estimate of the uncertainty which is particularly useful for reinforcement learning. This article explores the additional steps that are necessary to extend these methods to model multiple dialogue domains. We show that Gaussian process reinforcement learning is an elegant framework that naturally supports a range of methods, including prior knowledge, Bayesian committee machines and multi-agent learning, for facilitating extensible and adaptable dialogue systems.

Keywords: Dialogue systems; Reinforcement learning; Gaussian process

1. Introduction

Spoken dialogue systems allow humans to interact with machines using natural speech. As such, they have many benefits. By using speech as the primary communication medium, a computer interface can facilitate swift, human-like acquisition of information. In recent years, systems with speech interfaces have become ever more popular, as is evident from the rise of personal assistants such as Siri, Google Now, Cortana and Amazon Alexa. Statistical approaches to dialogue management have been shown to reduce design costs and provide superior performance to hand-crafted systems particularly in noisy environments (Young et al., 2013). Traditionally, spoken dialogue systems were built for limited domains described by an underlying *ontology*, which is essentially a structured representation of the database of entities that the dialogue system can talk about.

* Corresponding author at: University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK.
E-mail address: mg436@cam.ac.uk (M. Gašić).

http://dx.doi.org/10.1016/j.csl.2016.09.003

The semantic web is an effort to organise the large amount of information available on the Internet into a structure that can be more easily processed by a machine designed to perform reasoning on these data (Szeredi et al., 2014). *Knowledge graphs* are good examples of such structures. They typically consist of a set of triples, where each triple represents two entities connected by a specific relationship. Current knowledge graphs have millions of entities and billions of relations and are constantly growing. There has been a significant amount of work in spoken language understanding focused on exploiting knowledge graphs in order to improve coverage (Heck et al., 2013; Tür et al., 2012). More recently there have also been efforts to build statistical dialogue systems that operate on large knowledge graphs, but limited so far to the problem of belief tracking (Ma et al., 2015; Paul et al., 2015). In this article, we address the problem of decision-making in multi-domain dialogue systems. This a necessary step towards open-domain dialogue management. A previously proposed model for multi-domain dialogue management (Wang et al., 2015) assumes a dialogue expert for each domain and the central controller which decides to which dialogue expert to pass the control. The dialogue experts are rule-based and the central controller is optimised using reinforcement learning. A related work in Wang et al. (2015) proposes a domain independent feature representation of the dialogue state so that the dialogue policy can be applied to different domains. Here, we explore multi-domain dialogue management which retains a separate statistical model for each domain.

Moving from a limited domain dialogue system that operates on a relatively modest ontology size to an open domain dialogue system that can converse about anything in a very large knowledge graph is a non-trivial problem. An open domain dialogue system can be seen as a (large) set of limited domain dialogue systems. If each of them were trained separately then an operational system would require sufficient training data for each individual topic in the knowledge graph, which is simply not feasible. What is more likely is that there will be limited and varied data drawn from different domains. Over time, this data set will grow but there will always be topics within the graph which are rarely visited.

The key to statistical modelling of multi-domain dialogue systems is therefore the efficient reuse of data. Gaussian processes are a powerful method for efficient function estimation from sparse data. A Gaussian process is a Bayesian method which specifies a prior distribution over the unknown function and then given some observations estimates the posterior (Rasmussen and Williams, 2005). A Gaussian process prior consists of a *mean function* − which is what we expect the unknown function to look like before we have seen any data − and the *kernel function* which specifies the prior knowledge of the correlation of function values for different parts of the input space. For every input point, the kernel specifies the expected variation of where the function value will lie and once given some data, the kernel therefore defines the correlations between known and unknown function values. In that way, the known function values influence the regions where we do not have any data points. Also, for every input point the Gaussian process defines a Gaussian distribution over possible function values with mean and variance. When used inside a reinforcement learning framework, the variance can be used to guide exploration, avoiding the need to explore parts of the space where the Gaussian process is very certain. All this leads to very data efficient learning (Gašić and Young, 2014).

In this article, we explore how a Gaussian process-based reinforcement learning framework can be augmented to support multi-domain dialogue modelling focussing on three inter-related approaches. The first makes use of the Gaussian process prior. The idea is that where there is little training data available for a specific domain, a *generic* model can be used that has been trained on all available data. Then, when sufficient in-domain data become available, the generic model can serve as a prior to build a *specific* model for the given domain. This idea was first proposed in Gašić et al. (2015).

The second approach is based on a Bayesian committee machine (Tresp, 2000). The idea is that every domain or sub-domain is represented as a committee member. If each committee member is a Bayesian model, e.g. a Gaussian process, then the committee too is a Bayesian model, with mean and variance estimate. If a committee member is trained using limited data its estimates will carry a high uncertainty so the committee will rely on other more confident committee members, until it has seen enough training data. This method was proposed in Gašić et al. (2015). It is similar to Products of Gaussians which have previously been applied to problems such as speech recognition (Gales and Airey, 2006).

Finally, we extend the committee model to a multi-agent setting where committee members are seen as agents that collaboratively learn. This over-arching framework subsumes the first two approaches and provides a practical approach to on-line learning of dialogue decision policies for very large scale systems. It constitutes the primary contribution of this article.

The remainder of the paper is organised as follows. In Section 2, the use of Gaussian process-based reinforcement learning (GPRL) is briefly reviewed. The key advantage of GPRL in this context is that in addition to being data