Biologically Inspired Cognitive Architectures (2014) 9, 46–56

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bica

A layered architecture for probabilistic complex pattern recognition to detect user preferences

Michael Glodek [a,*], Thomas Geier [b], Susanne Biundo [b], Günther Palm [a]

[a] University of Ulm, Institute of Neural Information Processing, 89081 Ulm, Germany
[b] University of Ulm, Institute of Artificial Intelligence, 89081 Ulm, Germany

Received 15 April 2014; received in revised form 22 May 2014; accepted 14 June 2014

KEYWORDS
Layered architecture;
Markov models;
uncertainty

Abstract
The recognition of complex patterns is nowadays one of the most challenging tasks in machine learning, and it promises to be of great benefit for many applications, e.g. by allowing advanced human computer interaction to access the user's situative context. This work examines a layered architecture that operates on different temporal granularities to infer complex patterns of user preferences. Classical hidden Markov models (HMM), conditioned HMM (CHMM) and fuzzy CHMM (FCHMM) are compared to find the best configuration in the lower architecture layers. In the uppermost layer, a Markov logic network (MLN) is applied to infer the user preference in a probabilistic rule-based manner. For each layer a comprehensive evaluation is given. We provide empirical evidence showing that the layered architecture using FCHMM and MLN is well-suited to recognize patterns on different layers.

Introduction

Complex patterns are characterized by relational information derived from multiple sources including a temporal expansion (Gehrig et al., 2011; Yu & Ballard, 2004). They provide an important connecting factor to methods from symbolic artificial intelligence (AI) and interactive systems. Within this work we assume that complex patterns are composed of temporal sub-patterns that have a lower variability and are generally directly observable in the features extracted from the sensory (in contrast, complex patterns are assumed to be more abstract). Typical complex patterns are user emotions or dispositions which can be decomposed into behavioral cues (Scherer, Glodek, Layher et al., 2012; Vinciarelli, Pantic, Bourlard, & Pentland, 2008), or activities which can be decomposed into shorter actions (Gehrig et al., 2011; Nguyen, Phung, Venkatesh, & Bui, 2005).

* Corresponding author. Tel.: +49 500242424.
E-mail address: michael.glodek@uni-ulm.de (M. Glodek).

http://dx.doi.org/10.1016/j.bica.2014.06.003
2212-683X/

The ability to recognize complex patterns depends on various aspects: To ensure the robust detection of sub-patterns it is advisable to make use of multiple modalities, to handle uncertainty and to have suitable temporal fusion approaches. In addition, it is important to address the temporal nature of the problems by using sequential classifiers. On higher levels, a smooth shift to symbolic AI methods is inevitable, in particular because of the fact that the creation of data sets with a sufficient coverage of complex patterns is practically infeasible due to their large temporal variability. Symbolic AI techniques are often founded on spatial or temporal relations. However, in case of complex patterns comprising of multiple random variables, a suitable probabilistic logical framework is favorable. Because of their logical constraints, the derived complex classes can recover from wrongly recognized sub-patterns. It is even thinkable, that information extracted on the level of complex patterns can be back-propagated to the recognizers closer to the sensors to enhance the over-all performance.

In this work, we examine a layered architecture for complex pattern recognition. The study focuses on: the application of new kinds of hidden Markov models (HMM), the integration of symbolic and sub-symbolic information and the connection to AI methods. In the literature, there exists a number of related approaches. However, the focus is often set on solving a particular problem rather than evaluating a generic approach, such that a comparison to different methods is in general not straight forward. Especially because generally these data sets are not publicly available. Nonetheless, the state-of-art evolves and the achievements made so far are very encouraging (Geier, Biundo, Reuter, & Dietmayer, 2012; Glodek, Geier, Biundo, Schwenker, & Palm, 2013; Tenorth & Beetz, 2009; Wilson & Hendler, 1993; Wrede, Fritsch, Bauckhage, & Sagerer, 2004). Tran and Davis (2008) presented a system to enhance the surveillance of an outdoor parking lot utilizing a Markov logic network (MLN) (Richardson & Domingos, 2006). Kembhavi, Yeh, and Davis (2010) developed a system for scene understanding utilizing a MLN, integrating image analysis and reasoning. Oliver, Garg, and Horvitz (2004) proposed a multi-layered architecture to recognize office activities based on multiple sources. Gehrig et al. (2011) utilize a similar multi-layered approach to detect human intentions based on six activities and motion primitives. Domain knowledge is applied (by the means of ground truth data to learn transition probabilities) either to motion recognition, to activity recognition, or both.

The remainder of the work is organized as followed: Section ''Layered architecture'' describes the methodological aspects. In Section ''Experimental evaluation'' the experimental evaluation is presented. The conclusions are drawn in Section ''Conclusions''.

Layered architecture

The concept of layered architectures was first proposed by Oliver, Horvitz, and Garg (2002) in terms of layered hidden Markov models (LHMM) and is exemplified in Fig. 1 using two layers indexed by k. In the first layer ($k = 1$), a window is shifted over the features extracted from raw sensor data. The windowed data is passed to a set of HMM to recognize multiple sequential sub-patterns. The set of HMM outputs discrete classes which are then stored into the binary matrix $\mathbf{Z}^{(1)}$. The stream of stored recognition results is $\mathbf{Z}^{(k)} \in \mathbb{R}^{T \times I_k}$ where T represents the last time step available and I_k the number of classes of layer k. It is assumed, that the stored stream contains information about the classes from the higher levels. Therefore, in the subsequent layer ($k = 2$) a new window is shifted over the stream to extract a sequence of the recent history. The sequence is passed to a new set of discrete HMM to recognize the next set of (complex) patterns. Again the recognized classes are stored into a new binary matrix $\mathbf{Z}^{(2)}$. To increase the pattern complexity, the schema of the LHMM can be repeated.

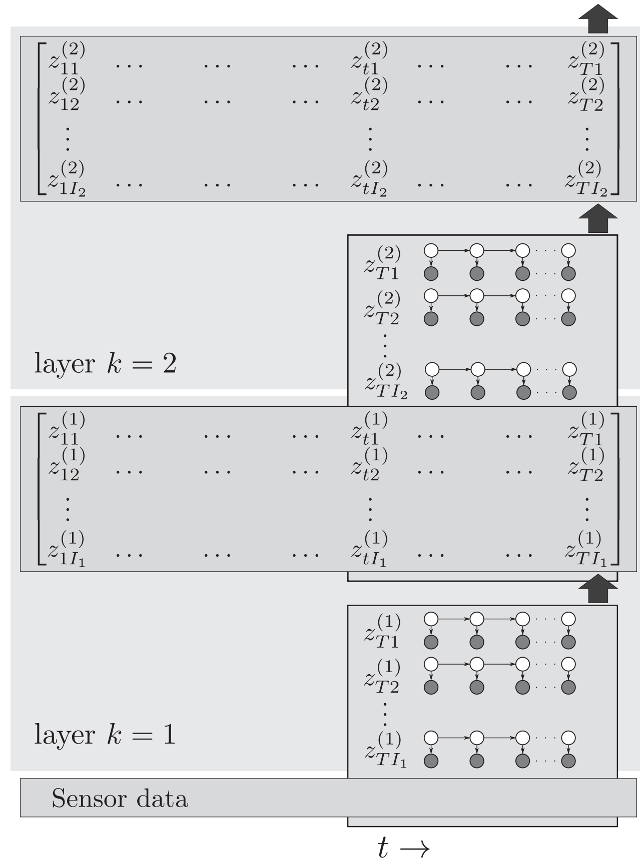

Fig. 1 Schematic drawing of the layered architecture concept. Windows are shifted over the sensory data associated with the lowest layer $k = 1$. The output of the HMM, operating on the windowed data and detecting sub-patterns, is collected into a stream which represents the input for the next layer $k = 2$. At the second layer an additional window is shifted over the compiled stream to extract data for the second set of HMM detecting the complex pattern. Figure adapted from Oliver et al. (2002).

Oliver et al. (2002) evaluated the LHMM in an office scenario to detect the desktop activities: ''phone conversation'', ''presentation'', ''face-to-face conversation'', ''user present'', ''engaged in some other activity'', ''distant conversation'' (outside the field of view); and ''nobody present''. The first layer detects classes utilizing the audio and video modalities. From the audio channel, the system recognizes the classes: ''human speech'', ''music'',