Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 47 (2018) 92–108

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

# Towards reasoning based representations: Deep Consistence Seeking Machine

Action editor: Simona Doboli

A. Lőrincz *, M. Csákvári, Á. Fóthi, Z. Á. Milacski, A. Sárkány, Z. Tősér

*Department of Software Technology and Methodology, Faculty of Informatics, Eötvös Loránd University, Pázmány Péter sétány 1/C, Budapest H-1117, Hungary*

Received 24 January 2017; received in revised form 8 August 2017; accepted 17 August 2017
Available online 30 August 2017

**Abstract**

Machine learning is making substantial progress in diverse applications. The success is mostly due to advances in deep learning. However, deep learning *can make mistakes* and its generalization abilities to new tasks are questionable. We ask when and how one can combine network outputs, when (i) details of the observations are evaluated by learned deep components and (ii) facts and rules are available. The Deep Consistence Seeking (DCS) machine seeks for consistent and deterministic event descriptions and improves the representation accordingly. The machine has an anomaly detection component that may trigger coherence seeking. Coherence seeking resolves conflicts between computational modules by preferring components with higher scores. We illustrate that context can help in correcting recognitions and in deriving training samples for self-training. We put these concepts into a general framework of cognition, by distinguishing creativity, rule extraction, verification, and symbol grounding. We demonstrate our approach in a driving scenario.


*Keywords:* Deep learning; Rule-based system; Recognition by components; Complexity; Communication; Episodic description

## 1. Introduction

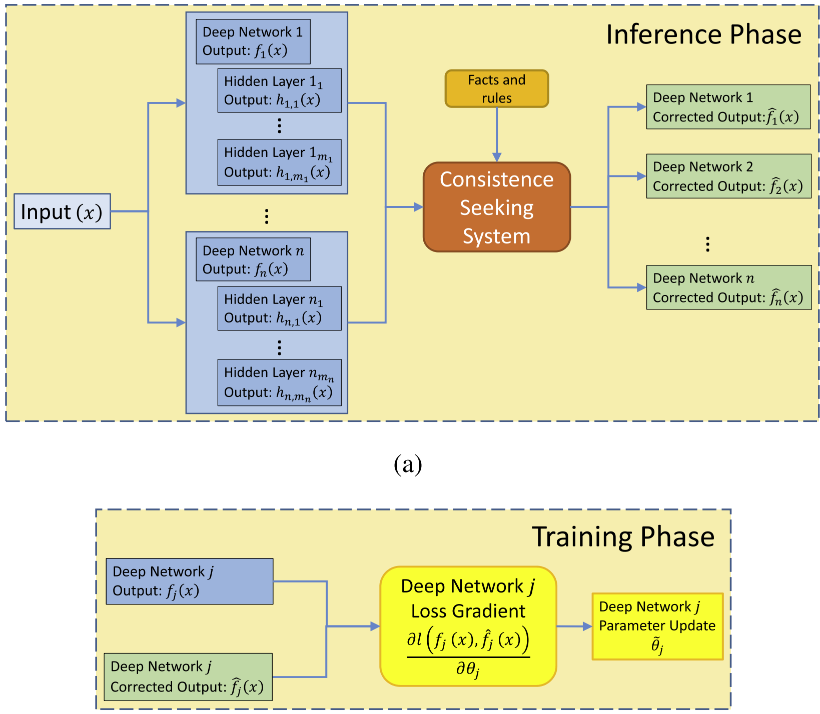
Machine learning is progressing quickly due to deep learning. The key tool for deep learning is crowdsourcing, i.e., the exploitation of human intelligence. Success stories demonstrate that superhuman performance can be reached this way (Schmidhuber, 2015). Still, the groundbreaking deep network approach seems limited as ‘each application requires years of focused research and careful unique construction’ (Study Panel, 2016). However, if we take a look at human information processing, for example, we learn that it has two basic routes: (i) holistic recognition (Tanaka & Gordon, 2011) and (ii) recognition by components (Biederman, 1987). These processing methodologies are competing and also complementing each other. Deep learning methods tend to favor end-to-end learning, which corresponds to holistic recognition and is thus fragile (Nguyen, Yosinski, & Clune, 2015; Sharif, Bhagavatula, Bauer, & Reiter, 2016).

We demonstrate that a traditional knowledge-based system is capable of combining and training several deep neural networks working on correlated components of a larger recognition problem. The knowledge-based system may include reasoning tools, differential equations, knowledge about the physics of the world, ontologies, among others. An important ingredient of our approach is the consistence seeking between *components that assume each other*.[1] From the point of view of machine learning, our procedure,

* Corresponding author.
*E-mail address:* lorincz@inf.elte.hu (A. Lőrincz).

[1] The task of learning of such components has been tackled recently by Lőrincz and Sárkány (2017).

http://dx.doi.org/10.1016/j.cogsys.2017.08.004
1389-0417/




(a)

(b)

Fig. 1. Schematic diagram of our Deep Consistence Seeking (DCS) machine. (a) Inference phase: the knowledge-based cognitive system settles inconsistencies between the components computed by deep neural networks using given rules, yielding corrected outputs. Typical deep networks have no consistence seeking, rule containing and corrected deep network components (for more details, see text). (b) Training phase: one may backpropagate errors between our predicted and newly obtained corrected outputs. Notation: $x$ is the input; $f_j(x)$ is the output of the $j$th deep network for input $x$; $\theta_j$ is the parameter vector of the $j$th deep network; $h_{j,m_j}(x)$ is the output of the $m$th hidden layer of the $j$th deep network for input $x$. Corrected quantities and updated parameters are denoted by hats and tilde, respectively.

which we call the *Deep Consistence Seeking* (DCS) machine, has two building blocks. The first one is discovering and resolving contradictions between the given components (Fig. 1(a)). The second part is fine-tuning the deep neural networks that produce the components: once inconsistencies are identified and corrected, new error-free samples are obtained that can be used for training (Fig. 1(b)). The corrected network outputs may then serve as features for other tasks (e.g., temporal segmentation or classification).

The first block tries to improve certainty. It assumes that the world is deterministic if sufficient information is available. It prefers components with high scores[2] and is willing to overwrite the internal processing of other deep components. If high score resolutions are achieved then they can be used for training. In turn, external supervision and component based supervision work in a similar manner. Behavioral feedback is the final reference. Under such conditions, a learning system should search for additional information if behavioral feedback is stochastic.

We put the mentioned machine learning methods into a general cognitive model that starts from considerations on computational complexity of solving problems and verifying them. We are concerned by the enigma that knowledge about nature collected by mankind in about 20,000 years can be passed to an individual in about 20 years or so. What took so long in these discoveries and why is it so easy to pass them? In trying to answer these questions, we highlight two categories of computational complexity: the hard do solve but easy to verify group, which are worth communicating; and the others that include the easy to solve but hard to verify group, being necessary for agreements and knowledge sharing. In our setup, the two classes correspond to exploring and exploiting relations between components, respectively. The learning of concepts or components and the relations between them may be seen as the extension of a symbolic system or the related rules.

We present our basic model in Section 2. We illustrate our approach on the State Farm Distracted Driver Scenario Kaggle benchmark in Section 3. We indicate that the deep components and the contradiction resolution together can improve performance and that unsupervised temporal segmentation can serve clustering and classification, too. The discussion (Section 4) considers the generality of the results. Using the algorithms in the benchmark we support our computational complexity thoughts and

[2] We use different input-output systems, some of them provide confidence values for their outputs, for example, our hand detector produces such information. We call such numbers scores. By contrast, our hand classifier outputs binary values.