

Abstract

This thesis develops new extensions to slow feature analysis (SFA) that solve supervised learning problems (e.g., classification and regression) on high-dimensional data (e.g., images) in an efficient, accurate, and principled way. This type of problems has been addressed by convolutional neural networks (CNNs) in the last decade with excellent results. However, additional approaches would be valuable, specially those that are conceptually novel and whose design can be justified theoretically.

SFA is an algorithm originally designed for unsupervised learning that extracts slow (i.e., temporally stable) features. Advantages of SFA include a strong theoretical foundation and that it might be intimately connected to learning in biological systems. One can apply SFA to high-dimensional data if it is implemented hierarchically, a technique called hierarchical SFA (HSFA). The extensions to SFA listed in the following allow the construction and training of deep HSFA networks, yielding competitive accuracy and efficiency.

Graph-based SFA (GSFA) is a supervised extension to SFA that introduces the concept of *training graph*, a structure in which the vertices are samples (e.g., images) and edges represent transitions between pairs of samples. Edges have weights that can be interpreted as desired output similarities of the corresponding samples. Compared to SFA, GSFA solves a more general optimization problem and considers many more transitions. Information about label (or class) similarities is encoded in the graph by the strength of the edge weights. Many training graphs are proposed to handle regression and classification problems. The efficacy of GSFA is demonstrated on a subproblem of face detection.

The *exact label learning* (ELL) method allows to compute training graphs where the slowest feature(s) one could extract would be equal to the label(s), if the feature space were unrestricted. In contrast to previously proposed graphs, the edge weights of the resulting ELL graphs are set precisely as needed, improving the label estimation accuracy. Moreover, ELL allows to learn multiple labels simultaneously using a single network, which is more efficient than learning the labels separately and often results in more robust features.

Hierarchical information-preserving GSFA (HiGSFA) improves the amount of label information propagated from the input to the top node in hierarchical GSFA (HGSFA). HiGSFA computes two types of features: slow features that maximize slowness, as usual, and reconstructive features that minimize an input reconstruction error, following an information-preservation goal. HiGSFA is evaluated on the problem of age estimation (along with gender and race) from facial photographs, where it yields a mean average error of 3.50 years, outperforming current state-of-the-art systems.

Among the proposed extensions, HiGSFA is the most promising. HiGSFA incorporates the other extensions and yields the best results, making this approach competitive, scalable, and robust. Moreover, HiGSFA is a versatile algorithm, allowing new technical applications and further principled extensions.

Kurzfassung der Dissertation

In dieser Doktorarbeit werden neue Erweiterungen des *Slow Feature Analysis* (SFA) Algorithmus vorgestellt. Diese lösen effizient und genau überwachte Lernprobleme des maschinellen Lernens (machine learning), wie Klassifikation und Regression, auf hochdimensionalen Daten (z.B. Bilder). Probleme dieser Klasse wurden in jüngerer Zeit hauptsächlich mit *Convolutional Neural Networks* behandelt, ein Ansatz der hervorragende Ergebnisse erzielt. Dennoch sind neue Lösungsansätze wünschenswert, insbesondere wenn sie auf neuen Konzepten basieren und Designentscheidungen auf fundierten theoretischen Grundlagen beruhen.

SFA ist ein Algorithmus der ursprünglich für das unüberwachte Lernen von langsamen (d.h. zeitlich stabilen) Merkmalen entwickelt wurde. Vorteile von SFA beinhalten ein im Detail ausgearbeitetes theoretisches Fundament sowie eine Verbindung zu Lernprozessen in biologischen Systemen. Eine hierarchische Implementierung von SFA (*hierachical SFA*, kurz: HSFA) erlaubt es, SFA auf hochdimensionale Daten anzuwenden. Die im folgenden genannten SFA-Varianten erlauben es, tiefe hierarchische SFA Netzwerke zu erstellen und zu trainieren. Diese erreichen Ergebnisse vergleichbar zu anderen Methoden bzgl. Genauigkeit und Effizienz.

Graph-based SFA (GSFA) ist eine SFA Variante für überwachtes Lernen, welche das Konzept des *training graph* einführt: eine Struktur in der die Knoten des Graphen die zu lernenden Datenpunkte und Kanten die Verbindungen zwischen diesen Datenpunkten repräsentieren. Die Gewichte der Kanten können als die gewünschte Ähnlichkeit zwischen zwei verbundenen Datenpunkten interpretiert werden. Im Vergleich zu SFA löst GSFA ein allgemeineres Optimierungsproblem und berücksichtigt eine wesentlich höhere Anzahl von Verbindungen zwischen einzelnen Datenpunkten. Die Information der Klassenähnlichkeit ist durch die Stärke der Verbindungsgewichte repräsentiert. In dieser Arbeit werden verschiedene Trainingsgraphen zum Lösen von Regressions- und Klassifikationsproblemen vorgestellt. Die Leistungsfähigkeit von GSFA wird anhand eines Unterproblems der Gesichtsdetektion demonstriert.

Der *exact label learning* (ELL) Algorithmus erlaubt es Graphen zu trainieren, für welche die ermittelten langsamen Merkmale den Klassenzugehörigkeiten entsprechen, falls der Merkmalsraum als unbeschränkt vorausgesetzt wird. Im Gegensatz zu den bisher vorgestellten Graphen werden die Verbindungsgewichte von ELL daher exakt vordefiniert, wodurch die Klassifikationsleistung des Netzwerks verbessert wird. Darüber hinaus erlaubt ELL das Lernen mehrerer Klassen mit nur einem einzigen Netzwerk. Dies ist nicht nur effizienter als ein separates Netzwerk für jede Klasse zu lernen, sondern führt außerdem zu robusteren Merkmalen.

Hierachical information-preserving GSFA (HiGSFA) vergrößert den durch das Netzwerk propagierenden Anteil der Information über die Klassenzugehörigkeit. HiGSFA extrahiert dazu zwei verschiedene Arten von Merkmalen: Langsame Merkmale, welche wie bisher die Langsamkeit maximieren, sowie Rekonstruktionsmerkmale, welche den Rekonstruktionsfehler minimieren und somit der Informationserhaltung dienen. Zur Evaluation der HiGSFA werden die Probleme der Alters- und Geschlechtsbestimmung und der ethnischen Zuordnung anhand von Porträtfotos herangezogen. Dabei erreicht HiGSFA im Falle des Ersteren einen durchschnittlichen Fehler von 3,5 Jahren und übertrifft mit diesem Ergebnis den bisherigen Stand der Technik.

Von den vorgestellten Varianten ist HiGSFA am vielversprechendsten. HiGSFA integriert die anderen aufgezählten Varianten und erzielt die besten Ergebnisse, was diesen Ansatz konkurrenzfähig, skalierbar und robust macht. Darüber hinaus bietet HiGSFA durch seine Vielseitigkeit die Möglichkeit neuer technischer Anwendungen sowie die Option auf weitere Varianten.