

A Factor Graph Approach to Signal Modelling, System Identification and Filtering

A dissertation submitted to the
Swiss Federal Institute of Technology, Zürich
for the degree of
Doctor of Technical Sciences

presented by

Sascha Korl

Dipl. Ing., TU Graz
born on August 8, 1974
citizen of the Republic of Austria

accepted on the recommendation of
Prof. Dr. Hans-Andrea Loeliger, examiner
Dr. Stefan Launer, co-examiner
Prof. Dr. Allen G. Lindgren, co-examiner

Hartung-Gorre Verlag, Konstanz, July 2005

Abstract

This thesis concerns model-based signal processing. In model-based signal processing a class of signals is described by a stochastic state-space model, in general with unknown parameters. The aim of signal estimation (filtering, denoising, parameter estimation, etc.) is to determine the 'best' signal (e.g. the most probable signal) out of that class given a set of observations.

Just a few years ago model-based signal processing was still limited to a restricted class of models. There are models with finite state-space (hidden Markov models) on the one hand and there are linear Gaussian state-space models (Kalman filter and related algorithms) on the other hand. For a multitude of applications those model types are insufficient.

Factor graphs open up new possibilities in this matter. First, factor graphs (and similar graphical models) provide a framework for the systematic and consistent derivation of classic model-based algorithms. Second, factor graphs permit and encourage the combination of different classic approaches for complex models with many unknown parameters; and third, factor graphs provide a framework for the systematic development of completely new algorithms.

A factor graph is used to represent the factorisation of the probability density function of the signal model. Inference is performed by passing messages along the edges of the graph. Messages can be interpreted as summaries of subgraphs, therefore the inference algorithm is called summary-propagation algorithm. We derive different message types justified by different representations of such summaries.

In a first step all messages in the graph are Gaussian. Several classic algorithms can be solely represented by Gaussian messages: Kalman filtering and smoothing, linear prediction and recursive least squares adaptive filters.

The more interesting and also more complicated case is when messages appear which are not Gaussian. Different techniques are proposed in this thesis to deal with such messages. For example, complicated messages can be represented with lists of samples of the exact message, which leads to particle-filter-type algorithms, or as gradients of the exact message, which leads to gradient descent (or hill climbing) methods.

The expectation maximisation (EM) algorithm is a powerful parameter estimation algorithm which has been used by many people in different applications. In this thesis we show how the EM algorithm can be stated as message passing on a factor graph. A simple message update rule is given, which allows the development of reusable building blocks. In varying the message update schedule, different new variants of the EM algorithm can be devised. Finally, a local message update rule arising from the combination of the ideas of summary-propagation and EM is given.

The utilisation of the proposed techniques is demonstrated by means of the autoregressive model with unknown coefficients, unknown input noise variance and unknown observation noise variance.

Keywords: Graphical models, factor graphs, summary-propagation algorithm, belief propagation, message passing, expectation maximisation, signal modelling, system identification, autoregressive model, Kalman filter.

Kurzfassung

Die vorliegende Arbeit befasst sich mit modellbasierter Signalverarbeitung. Bei der modellbasierten Signalverarbeitung wird eine Klasse von Signalen durch ein stochastisches Zustandsraummodell, im Allgemeinen mit unbekanntem Parametern, beschrieben. Das Ziel der Signalschätzung (Filterung, Entrauschung, Parameterschätzung, etc.) ist es das 'beste' Signal aus der modellierten Signalklasse zu ermitteln, wobei alle verfügbaren Beobachtungen berücksichtigt werden sollten.

Noch vor wenigen Jahren war die modellbasierte Signalverarbeitung auf wenige Typen von Modellen beschränkt. Im Wesentlichen gab es einerseits Modelle mit endlichem Zustandsraum (Hidden Markov Modelle) und andererseits lineare Modelle mit Gauss'schem Rauschen (Kalmanfilter und verwandte Algorithmen). Für eine Vielzahl von Anwendungen sind diese Modelltypen aber nicht ausreichend.

Faktorgraphen eröffnen hier neue Horizonte. Erstens bieten Faktorgraphen (und ähnliche graphische Modelle) einen Rahmen für die systematische und einheitliche Herleitung klassischer modellbasierter Signalverarbeitungs-Algorithmen; zweitens ermöglichen und ermuntern Faktorgraphen zur kombinierten Verwendung verschiedener klassischer Verfahren für komplexe Modelle mit vielen unbekanntem Parametern und drittens bieten Faktorgraphen einen Rahmen zur systematischen Entwicklung gänzlich neuer Verfahren.

Ein Faktorgraph wird verwendet um eine Faktorisierung der Wahrscheinlichkeitsdichtefunktion des Signalmodells zu präsentieren. Statistische Inferenz erfolgt durch das Versenden von Nachrichten entlang der Kanten des Graphen. Diese Nachrichten können als Zusammenfassungen

von Teilen des Graphen gesehen werden, daher der Name Summary-Propagation-Algorithmus. Es werden verschiedene Nachrichtentypen abgeleitet, welche sich durch verschiedene Repräsentationen dieser Zusammenfassungen ergeben.

In einer ersten Betrachtung sind alle Nachrichten gaussförmig. Der Kalmanfilter und -smoother, lineare Prädiktion und recursive least squares adaptive Filter können mit ausschliesslich gaussförmigen Nachrichten dargestellt werden.

Der interessantere und kompliziertere Fall tritt ein, wenn Nachrichten entstehen, die nicht mehr als Gauss-Funktion dargestellt werden können. Verschiedene Methoden mit solchen Nachrichten umzugehen werden in dieser Arbeit behandelt. Zum Beispiel können komplizierte Nachrichten als Liste von Samples der exakten Nachricht dargestellt werden, was zu Particle-Filter Algorithmen führt. Oder der Gradient wird zur Darstellung der Nachricht herangezogen, was zu gradient descent (oder hill climbing) Methoden führt.

Der Expectation-Maximisation-Algorithmus (EM) ist ein leistungsfähiger Algorithmus zur Parameterschätzung und wird von Vielen in verschiedenen Anwendungen eingesetzt. In dieser Dissertation wird gezeigt, wie der EM-Algorithmus als Message-Passing am Faktorgraphen dargestellt werden kann. Eine einfache Nachrichten-Aufdatierungsregel erlaubt die Entwicklung von wiederverwendbaren Bausteinen. Neue Varianten des EM-Algorithmus erhält man durch Variation der Reihenfolge der Nachrichten-Aktualisierung. Zum Schluss wird eine lokale Nachrichten-Aufdatierungsregel angegeben, welche Ideen von EM und Summary-Propagation kombiniert.

Die Anwendung der vorgestellten Methoden wird anhand des autoregressiven Modells mit unbekanntem Koeffizienten, unbekanntem Eingangs- und Beobachtungsräuschen demonstriert.

Stichworte: Graphische Modelle, Faktorgraphen, Summary-Propagation Algorithms, Belief Propagation, Message Passing, Expectation Maximisation, Signalmodellierung, System-Identifikation, Autoregressives Modell, Kalmanfilter.