Übung zur Vorlesung

Maschinelles Lernen 2

Sommersemester 2011

Abteilung Maschinelles Lernen
Institut für Softwaretechnik und
theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Blatt 3

Abgabe bis Montag, **9. Mai 2011**, 12:00 Uhr, Briefkasten bei Raum FR6061

1. **(Breitbandige BSS für Faltungsmixturen, generisches Koeffizientenupdate basierend auf natürlichen Gradienten, 12 Punkte)**

   Wie in der Vorlesung dargestellt, lassen sich zahlreiche unterschiedliche Aufgabenstellungen in der Signalverarbeitung durch iterative Adaption von MIMO-FIR Filterkoeffizienten mittels Updates der folgenden Struktur (sog. *natürlicher Gradient*) lösen:

$$\Delta\check{\mathbf{W}}^{\ell}(m) = \sum_{i=0}^{\infty} \beta(i,m)\, \mathcal{SC} \left\{ \mathbf{W}^{\ell-1}(i) \left[\underbrace{\frac{1}{N} \sum_{j=iN_L}^{iN_L+N-1} \mathbf{y}(j) \mathbf{\Phi}_{s,PD}^{\mathrm{T}}(\mathbf{y}(j))}_{=:\mathbf{R}_{\mathbf{y}\mathbf{\Phi}(\mathbf{y})}} - \mathbf{I} \right] \right\} \tag{1}$$

   mit der *multivariaten Score-Funktion*

$$\mathbf{\Phi}_{s,PD}(\mathbf{y}(j)) = -\frac{\partial \log \hat{p}_{s,PD}(\mathbf{y}(j))}{\partial \mathbf{y}(j)}. \tag{2}$$

   Zeige, dass sich diese generische Darstellung für den praktisch wichtigen Fall, gegeben durch die Forderungen

   - BSS-Anwendung: Wechselseitige statistische Unabhängigkeit zwischen den Ausgangskanälen (und dadurch *Partitionierung* des PD-variaten Vektors $\mathbf{\Phi}_{s,PD}$ der Score-Funktion)
   - Annahme von sogenannten *Spherically Invariant Random Processes* (SIRPs) für die einzelnen Auskangskanäle y_p, $p = 1, \ldots, P$ mit der allgemeinen Form

$$\hat{p}_{y_p,D}(\mathbf{y}_p(j)) = \frac{1}{\sqrt{\pi^D \det(\mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}(i))}} f_{p,D}\left(\mathbf{y}_p^{\mathrm{T}}(j) \mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}^{-1}(i) \mathbf{y}_p(j)\right), \tag{3}$$

   folgendermassen spezialisieren lässt:

$$\Delta\check{\mathbf{W}}^{\ell}(m) = \sum_{i=0}^{\infty} \beta(i,m)\, \mathcal{SC} \left\{ \mathbf{W}^{\ell-1}(i) \left[\mathbf{R}_{\mathbf{y}\phi(\mathbf{y})}(i)\, \mathrm{bdiag}^{-1}\left\{\mathbf{R}_{\mathbf{y}\mathbf{y}}(i)\right\} - \mathbf{I} \right] \right\}, \tag{4}$$

   wobei $\mathbf{R}_{\mathbf{y}\phi(\mathbf{y})}(i)$ aus den folgenden Untermatrizen besteht:

$$\mathbf{R}_{\mathbf{y}_p\phi(\mathbf{y}_q)}(i) := \frac{2}{N} \sum_{j=iN_L}^{iN_L+N-1} \mathbf{y}_p(j) \phi_{y_q,D}(u_q(j)) \mathbf{y}_q^{\mathrm{T}}(j). \tag{5}$$

   mit $\phi_{y_p,D}(u_p) := -\frac{1}{f_{p,D}(u_p)} \frac{\partial f_{p,D}(u_p)}{\partial u_p}$ und $u_p = \mathbf{y}_p^{\mathrm{T}} \mathbf{R}_{\mathbf{y}_p\mathbf{y}_p}^{-1} \mathbf{y}_p$.

2. **(Näherungsweise BSS für Faltungsmixturen im Frequenzbereich, 7 Punkte)** Auf der Webseite findet man zwei wav-Dateien, welche real aufgezeichnete Mischungen (Mikrophonsignale) von zwei menschlichen Sprechern in einem Raum enthalten (Nachhallzeit $T_{60} \approx$ 50ms, entspricht etwa einem Innenraum eines PKWs), sowie einer *näherungsweisen* Realisierung in Matlab für BSS im Frequenzbereich (DFT-Bereich) und mit Annahme von approximierten Laplace-Dichten.

   Mit der Markierung "!!!!!!! " findet man im Matlab-Skript zwei Varianten, die (a) eine völlig voneinander unabhängige Behandlung der Frequenzkomponenten definieren, oder (b) eine Kopplung, d.h. gemeinsame (breitbandige) Behandlung aller Frequenzkomponenten bewirken.

   Für beide Varianten soll jeweils das Matlab-Skript ausgeführt werden. Anschliessend sollen die mitgelieferten Mischsignale `x1.wav`, `x2.wav`, sowie die jeweils sich ergebenden Ausgangssignale `y1.wav`, `y2.wav` abgehört werden. Es sind die folgenden Fragen zu beantworten (keine ausführliche Mathematik nötig):

   - Wie sind die Ergebnisse der beiden Varianten zu interpretieren? Warum unterscheiden sich die Separationsergebnisse beider Varianten deutlich?
   - Wie kommt die Kopplung der Frequenzkomponenten in dieser Realisierung zustande? (Man kann hier z.B. anhand der Analogie zum SIRPs-basierten (exakten) Zeitbereichsalgorithmus aus Aufgabe 1 argumentieren.)
   - Worin würde selbst bei perfekter Kopplung der DFT-Komponenten die Näherung gegenüber einer exakten Zeitbereichsimplementierung bestehen?

3. **(Ideale Separationslösung / Anwendung zur Lokalisierung von simultan aktiven Schallquellen, 11 Punkte)**

   In dieser Aufgabe sollen anhand der BSS-Realisierung aus Aufgabe 2 die ursprünglichen Quellenpositionen im Raum geschätzt werden. Gegeben sind lediglich die Samplingrate der Mikrophonsignale von 16kHz, der räumliche Abstand zwischen den beiden Mikrophonen von 12.6cm, sowie die Schallgeschwindigkeit in Luft von $c = 334\text{m/s}$.

   (a) Man fertige zunächst eine Skizze an, die die Kaskade aus MIMO-Mischsystem und MIMO-Entmischsystem enthält. Die einzelnen Pfade sollen jeweils korrekt beschriftet werden ($h_{11}, \ldots$ bzw. $w_{11}, \ldots$). Ausserdem sollen die Einfallswinkel Θ_1 und Θ_2 zwischen den Quellensignalen und der zum Array senkrechten Ebene eingezeichnet werden.

   (b) Mittels der *idealen Separationslösung* und des mit dem Code aus Aufgabe 2 geschätzten Entmischsystems soll auf die Pfade des Mischsystems zurückgeschlossen werden, und daraus für beide Quellen jeweils die Laufzeitverzögerungen (Maxima der Absolutbeträge der Impulsantworten) zwischen den beiden Mikrophonen bestimmt werden (*Time-Differences of Arrival, TDOAs*). Aus Gründen der Genauigkeit sollen die identifizierten Impulsantworten vor der Ermittlung der Maxima um einen Faktor 10 hochgesampelt werden (Matlab-Befehl `resample`).

   (c) Mittels *Fernfeldnäherung* (d.h., Annahme von ebenen Wellenfronten) können aus den TDOAs die beiden Winkel Θ_1 und Θ_2 berechnet werden (siehe obige Angaben zu Mikrophonabstand und Schallgeschwindigkeit).

Für Fragen zum Übungsblatt bitte in der Google Group http://groups.google.com/group/ml-tu registrieren und die Fragen an die Mailingliste stellen.