Multimodale Registrierung mit effizienten Lernverfahren für neuronale Netze

Torsten Rohlfing, Jürgen Beier, Oliver Schulte, Norbert Hosten und Roland Felix

Virchow-Klinikum, Medizinische Fakultät der Humboldt-Universität zu Berlin, Strahlen- und Poliklinik, Augustenburger Platz 1, 13353 Berlin
Email: torsten@ukrv.de

Zusammenfassung. Ausgangspunkt ist die mathematische Äquivalenz von voxelbasierter Registrierung und dem Training neuronaler Netze als multidimensionales Optimierungsproblem. Die Lernverfahren Quickprop und Rprop wurden implementiert, modifiziert und als Optimierungsverfahren für ein voxelbasiertes Ähnlichkeitsmaß verwendet. Tests mit realen Daten zeigten eine deutliche Überlegenheit beider Verfahren gegenüber einfachem Gradientenabstieg, wobei sich ein Vorteil zugunsten von Rprop andeutete. Der Vergleich mit einem konjugierte-Gradienten-Verfahren ergab keine signifikanten Unterschiede bezüglich Effizienz und Stabilität.

Schlüsselwörter: Voxelbasierte Registrierung, Optimierung, adaptive neuronale Lernverfahren

1 Einleitung

Die Registrierung zweier medizinischer Bilddatensätze (Referenz- und Modelldatensatz) erfordert das Finden einer geometrischen Transformation, welche die Koordinaten des Modells in die der Referenz überführt.

Durch Definition eines Maßes S für die erreichte Qualität der Registrierung und Festlegung der zulässigen Transformationen Ω ergibt sich in natürlicher Weise folgendes Optimierungsproblem: Unter allen Transformationen $\boldsymbol{\omega} \in \Omega$ ist diejenige zu finden, welche angewendet auf den Modelldatensatz bezüglich des (ortsfesten) Referenzdatensatzes die Funktion $S(\boldsymbol{\omega})$ maximiert.

Dabei ist prinzipiell unerheblich, welche Transformationen (affin, elastisch) zugelassen sind und wie S definiert ist (Kreuzkorrelation, mittleres Abstandsquadrat, gewichtete Varianzen, mutual information).

Analog läßt sich das Lernproblem bei neuronalen Assoziativspeichern (feed forward Netze) als Optimierungsproblem formulieren. Die Optimierungsfunktion ist in diesem Fall der (zu minimierende) Lernfehler des Netzes. Ebenso wie daher gängige Optimierungsverfahren (downhill-Suche, konjugierte Gradienten, simulated annealing etc.) zum Trainieren neuronaler Netze verwendet werden, können umgekehrt effiziente Lernverfahren auch allgemeinere Optimierungsprobleme lösen.

Gefördert von der Deutschen Forschungsgemeinschaft, GRK 331–1/97.

2 Material und Methoden

2.1 Bildakquisition

Für den Vergleich der Optimierungsverfahren wurden dreidimensionale Computer- (CT) und Magnetresonanztomographien (MRT) aus dem Schädelbereich von 7 Tumorpatienten verwendet.

Für die CT-Akquisition wurden folgende Geräte und Einstellungen verwendet: Somatom Plus S (Siemens AG, Erlangen), Spiraltechnik, axiale Schichtung, Schichtabstand 5mm, Schichtauflösung 0,57–0,70mm, Rekonstruktionsmatrix 512×512 Pixel.

Die MRT-Aufnahmen wurden auf einem Magnetom (Siemens AG, Erlangen) mit folgenden Parametern durchgeführt: T1-gewichtete axiale Schichten, Schichtabstand 5–6mm, Pixelgröße 0,90–1,17mm, Matrixgröße 256×256 Pixel.

2.2 Iterative Optimierungsverfahren

Die Registrierung der Datensätze mittels starrer-Körper-Transformation erfolgte durch Maximierung des entropiebasierten Ähnlichkeitsmaßes "mutual information" nach Wells et al. [1]. Aufbauend auf dem kommerziellen Bildverarbeitungssystem AVS/Express (AVS/UNIRAS, Erkrath) wurden eigene Softwarekomponenten implementiert, welche die Optimierungsverfahren Gradientenabstieg, Quickprop, Rprop und konjugierte Gradienten realisieren.

Beginnend mit einer Ausgangsnäherung $\omega^{(0)}$ (initiale Registrierung) wird in einem iterativen Prozeß jeweils aus einer Näherungslösung $\boldsymbol{\omega}^{(i)}$ die nächste Approximation $\boldsymbol{\omega}^{(i+1)}$ durch einen Schritt $\Delta\boldsymbol{\omega}^{(i)}$ berechnet:

$$\boldsymbol{\omega}^{(i+1)} = \boldsymbol{\omega}^{(i)} + \Delta\boldsymbol{\omega}^{(i)}. \tag{1}$$

Gradientenabstieg (GD). Wird $\Delta\boldsymbol{\omega}^{(i)}$ in (1) immer in Richtung des lokalen Gradienten von S gewählt als

$$\Delta\boldsymbol{\omega}^{(i)} = \delta \frac{d}{d\boldsymbol{\omega}} S(\boldsymbol{\omega}^{(i)}), \tag{2}$$

erhält man einen reinen Gradientenabstieg. Ein freier Parameter δ regelt hierbei die Schrittweite. Angewendet auf das Training neuronaler Netze ist dies äquivalent zum Backpropagation-Algorithmus. Rumelhart et al. [2] führen ein *Trägheitsmoment* m ein und modifizieren (2) zu

$$\Delta\boldsymbol{\omega}^{(i)} = \delta \frac{d}{d\boldsymbol{\omega}} S(\boldsymbol{\omega}^{(i)}) + m\Delta\boldsymbol{\omega}^{(i-1)}. \tag{3}$$

Quickprop. Das von Fahlman entwickelte *Quickprop*-Verfahren [3] basiert auf der Annahme, die zu optimierende Funktion verhalte sich in jeder Dimension unabhängig von den anderen Parametern lokal quadratisch. Die mehrdimensionale Verallgemeinerung des Newton-Verfahrens führt für diese Approximation zu

$$\Delta\omega_j^{(i)} = \frac{\frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i)})}{\frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i)}) - \frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i-1)})} \Delta\omega_j^{(i-1)}. \tag{4}$$

In der Praxis wird die Schrittweite auf ein ν-faches des jeweils letzten Schrittes beschränkt, um die numerische Stabilität der Iteration zu gewährleisten. Zusätzlich wird jedem Schritt ein mit ε gewichteter Gradiententerm überlagert.

Resilient Backpropagation (Rprop). Bei dem von Riedmiller und Braun [4] unter dem Namen "resilient backpropagation" vorgestellten Verfahren entfällt die Abhängigkeit der Schrittweite vom Betrag des Gradienten. Lediglich seine Richtung wird für den nächsten Schritt in Betracht gezogen, während die Schrittweite unabhängig vom lokalen Gradienten bestimmt wird:

$$\Delta\omega_j^{(i)} = \begin{cases} -\Delta_j^{(i)} & \text{falls } \frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i)}) < 0 \\ +\Delta_j^{(i)} & \text{falls } \frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i)}) > 0 \\ 0 & \text{falls } \frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i)}) = 0 \end{cases} \tag{5}$$

Nach Komponenten getrennt wird die Schrittweite adaptiv um einen Faktor η^- verringert oder um η^+ vergrößert, wobei gilt $0 < \eta^- < 1 < \eta^+$:

$$\Delta_j^{(i)} = \begin{cases} \eta^+ \cdot \Delta_j^{(i-1)} & \text{falls } \frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i-1)}) \frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i)}) > 0, \\ \eta^- \cdot \Delta_j^{(i-1)} & \text{falls } \frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i-1)}) \frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i)}) < 0, \\ \Delta_j^{(i-1)} & \text{falls } \frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i-1)}) \frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i)}) = 0. \end{cases} \tag{6}$$

Conjugate Gradients (CG). Als Referenz wurde ein klassischer Optimierer nach der Methode der konjugierten Gradienten in der Variante von Polak und Ribiere eingesetzt (vgl. Press et al. [5]).

2.3 Eigene Anpassungen der Algorithmen

Für die spezielle Situation wurden die oben beschriebenen Standardverfahren modifiziert und an die Besonderheiten der Registrierung angepaßt.

Gradientenabstieg. Um die Konvergenz zu verbessern, erfolgte nach einer geeignet gewählten maximalen Anzahl von Verschlechterungen ein erzwungenes Rücksetzen (backtracking) auf die beste bisher gefundene Transformation.

Resilient Backpropagation. Abweichend von (6) erfolgte auch bei $\frac{\partial}{\partial\omega_j} S(\boldsymbol{\omega}^{(i)}) = 0$ ein Verringern der Schrittweite um η^- mit dem Ziel, die Zerstörung von Stabilität einer Variablen durch plötzliche große Schritte zu verhindern.

Conjugate Gradients. Die Optimierung von Translation und Rotation geschah zwar parallel, doch mathematisch unabhängig voneinander, da durch unterschiedliche Basiseinheiten (Millimeter vs. Grad) die partiellen Gradienten nicht kompatibel sind.

Initiale Translation. Alle Optimierungsverfahren wurden sowohl mit als auch ohne eine initiale Translation getestet, durch die vor der eigentlichen Registrierung die Mittelpunkte beider Datensätze zur Deckung gebracht wurden.

3 Ergebnisse

Im Durchschnitt konvergierten sowohl Rprop als auch Quickprop zuverlässiger und schneller als einfacher Gradientenabstieg (vgl. Tab. 1). Das Konvergenzverhalten der konjugierte-Gradienten-Methode war den lokalen Optimierern dabei nur knapp überlegen.

In insgesamt 20/56 Fällen schlug die Registrierung fehl (GD 8/14, Rprop 5/14, Quickprop 6/14, CG 1/14). Initiale Translation des Modelldatensatzes ermöglichte in 7/13 zuvor erfolglosen Fällen das Finden einer akzeptablen Registrierung. In einem Fall (Quickprop bei Patient #2) wurde eine vorher erreichte Lösung mit initialer Translation nicht mehr gefunden.

Initiale Translation beschleunigte die Konvergenz in 16/28 Fällen, in 4/28 Fällen konvergierten die Iterationen ohne sie besser. War die Registrierung sowohl mit als auch ohne initiale Translation erfolgreich (14/28 Fälle), wurde die Qualität der Registrierung in 7/14 Fällen durch die initiale Translation verbessert (Tab. 2).

4 Diskussion

Die insgesamt große Zahl nicht gefundener Registrierungen erklärt sich durch die teilweise stark unterschiedlichen Bildausschnitte der einzelnen Modalitäten. Daran vermag die verwendete initiale Registrierung kaum etwas zu ändern.

	Patient						
Optimierer	1	2	3	4	5	6	7
GD	–/–	–/–	30/20	–/–	23/19	20/21	–/–
Rprop	–/–	16/14	9/11	–/8	–/11	10/8	–/6
Quickprop	–/–	18/–	14/11	–/10	–/12	11/9	–/8
CG†	–/11	5/5	3/3	7/2	14/6	5/1	6/7

Tabelle 1. Iterationen bis zur Konvergenz ohne (erster Wert) und mit initialer Translation (zweiter Wert). Ist kein Eintrag vorhanden, lieferte die jeweilige Iteration kein brauchbares Ergebnis. †Die Anzahl der Iterationen beim konjugierte-Gradienten-Verfahren ist nicht mit den übrigen Angaben vergleichbar, da pro Iteration eine projizierte Optimierung erforderlich ist.

	Patient						
Optimierer	1	2	3	4	5	6	7
GD	–/–	–/–	0.93/0.94	–/–	0.88/0.89	0.88/0.89	–/–
Rprop	–/–	0.98/0.96	0.85/0.86	–/1.03	–/0.88	0.83/0.87	–/0.95
Quickprop	–/–	0.98/–	0.87/0.87	–/1.06	–/0.88	0.86/0.88	–/0.93
CG	–/0.76	0.99/0.99	0.85/0.85	1.05/1.05	0.86/0.88	0.85/0.85	0.93/0.93

Tabelle 2. Erreichte Registrierungsqualität entsprechend Ähnlichkeitsmaß "mutual information" (Wertebereich $\mathbb{R}^+$, höhere Werte zeigen bessere Qualität an) nach Wells et al. [1]. Der jeweils zweite Wert wurde mit initialer Translation berechnet.

Hiervon abgesehen zeigten die durchgeführten Tests, daß sowohl Rprop als auch Quickprop dem Gradientenabstieg überlegen sind und dem konjugierte-Gradienten-Verfahren hinsichtlich Effizienz und Stabilität kaum nachstehen. Dabei deutete sich eine leichte Überlegenheit von Rprop gegenüber Quickprop an, was auch Trainings-Benchmarks für neuronale Netze belegen [4].

Daß allerdings die Optimierung eines Ähnlichkeitsmaßes nicht notwendigerweise identisch mit dem Auffinden einer geeigneten Registrierungstransformation ist, zeigen insbesondere die Resultate für Patient #6 mit initialer Translation. Das vom Gradientenabstieg erreichte Ergebnis lag über dem der anderen Verfahren, obwohl nach visueller Beurteilung den von letzteren gelieferten Transformationen eindeutig der Vorzug zu geben wäre.

Insofern ist der Einwand berechtigt, ein abstraktes Ähnlichkeitsmaß (Tab. 2) sei kein geeignetes Kriterium für den Vergleich von Registrierungsverfahren. Dem ist zu entgegnen, daß das Fehlen eines globalen Maximums bei der korrekten Transformation ein Mangel des Ähnlichkeitsmaßes und nicht etwa des Optimierungsverfahrens ist. Dessen Leistungsfähigkeit kann daher durchaus anhand der angegebenen Daten bewertet werden, weswegen auf einen Vergleich mit manuell oder mittels Landmarken gefundenen Registrierungen verzichtet wurde.

Literatur

1. Wells WM, Viola P, Atsumi H, Nakajima S, Kikinis R: Multi-Modal Volume Registration by Maximization of Mutual Information. In: *Medical Image Analysis*, 1(1), 1996
2. Rumelhart DE, Hinton GE, Williams RJ: Learning internal representations by error propagation. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1:318–362. MIT Press, Cambridge, MA, 1986
3. Fahlman SE: An Empirical Study of Learning Speed in Back-Propagation Networks. Technical Report CMU-CS-88-162, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1988
4. Riedmiller M, Braun H: A direct adaptive method for faster backpropagation algorithm learning: The RPROP algorithm. In: *Proc. of IEEE Int. Conf. on Neural Networks (ICNN)*, 586–591, San Francisco, 1993
5. Press WH, Teukolsky SA, Vetterling WT, Flannery BP: *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1988