Redes de Hopfield Quaterniônicas

Aluno: Fidelis Zanetti de Castro, Ms.C.

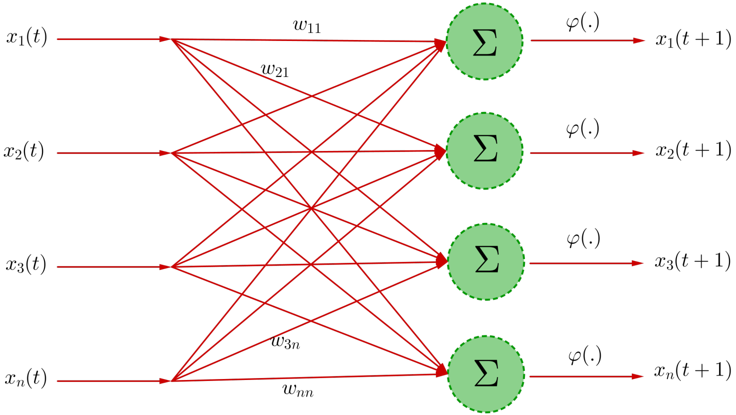
Orientador: Marcos Eduardo Ribeiro do Valle Mesquita, Ph.D.

IMECC – UNICAMP

20 de março de 2015

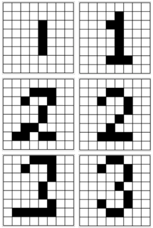
A rede de Hopfield Clássica

- A rede neural de Hopfield é um caso particular de rede recorrente, em que o espaço de estados é discreto (rede neural de primeira geração).



A rede de Hopfield Clássica

- Ela pode ser utilizada para implementar uma memória associativa não linear (endereçada por conteúdo).
- Formalmente, uma memória associativa (AM) corresponde a uma aplicação $\mathcal{M}$, chamada *aplicação associativa*, tal que $\mathcal{M}(\mathbf{u}^\xi) = \mathbf{u}^\xi$ para todo $\xi = 1, \ldots, p$. Além disso, deseja-se que uma AM exiba alguma tolerância a ruído no sentido que $\mathcal{M}(\mathbf{x}) = \mathbf{u}^\xi$ sempre que $\mathbf{x}$ corresponder a uma versão corrompida ou distorcida de $\mathbf{u}^\xi$.



A rede de Hopfield Clássica

- O modelo da rede neural de Hopfield utiliza como unidade de proces - samento básica o neurônio de MCCULLOCH & PITTS (1943).
- Em geral, neste tipo de rede não há auto-alimentação, ou seja, a saída de um neurônio alimenta todos os outros neurônios, exceto ele mesmo.
- Em termos matemáticos, uma rede de Hopfield com n neurônios e sem auto-alimentação define recursivamente uma sequência de vetores binários $\mathbf{x}(0), \mathbf{x}(1), \mathbf{x}(2), \ldots$ em $\{-1, +1\}^n$ por meio da seguinte equação:

$$x_j(t+1) = sgn\left(\sum_{k \neq j} w_{jk} x_k(t)\right), \quad \forall j = 1, \ldots, n$$

- Na equação acima, $\mathbf{x}(0) = [x_1, \ldots, x_n]^T \in \{-1, +1\}^n$ é um estado inicial e $t = 0, 1, \ldots$

A rede de Hopfield Clássica

- O número w_{jk} representa o peso da k-ésima conexão sináptica do j-ésimo neurônio.
- Na rede de Hopfield, a função sinal é avaliada no número $v_j(t) = \sum_{k \neq j} w_{jk} x_k(t)$, com $j = 1, \ldots, n$, que é denominado *potencial de ativação do j-ésimo neurônio no instante t.*
- Tem-se o seguinte:

$$sgn(v_j(t)) = \begin{cases} +1, & v_j(t) > 0, \\ x_j(t), & v_j(t) = 0, \\ -1, & v_j(t) < 0. \end{cases} \tag{1}$$

A rede de Hopfield Clássica

- Quando a rede de Hopfield é usada para implementar uma memória associativa, as conexões sinápticas w_{jk} são originalmente determinadas da seguinte forma: Dado um conjunto de vetores $\{\mathbf{u}^1, \ldots, \mathbf{u}^p\}$, em que $\mathbf{u}^\xi = [u_1^\xi, \ldots, u_n^\xi]^T$ com $u_i^\xi \in \{-1, +1\}^n$, define-se

$$w_{jk} = \begin{cases} \sum_{\xi=1}^{p} u_j^\xi u_k^\xi, & i \neq j \\ 0, & i = j, \end{cases} \tag{2}$$

  para todo $i, j = 1, \ldots, n$.
- A regra definida em (2) é referida como *armazenamento por correlação* ou *aprendizado de Hebb* sem auto-alimentação.
- A seguinte equação fornece uma função energia para a rede de Hopfield:

$$E(x) = -\frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} x_j w_{jk} x_k, \quad \forall \mathbf{x} \in \{-1, +1\}^n. \tag{3}$$

Conceitos básicos sobre os quatérnios

- Os quatérnios foram introduzidos pelo irlandês *William Rowan Hamilton* na sua obra *Álgebra dos quatérnios* em meados do século XIX.
- São números que podem ser escritos na forma $q = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$ onde $\mathbf{i}$, $\mathbf{j}$ e $\mathbf{k}$ são três tipos de números imaginários (unidades hiperimaginárias) que satisfazem $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$.
- A álgebra dos quatérnios é a única álgebra associativa e divisional que pode ser definida ao lado da álgebra dos reais e dos complexos.
- Um quatérnio pode ser visto como uma quádrupla de números reais, isto é, $q = (q_0, q_1, q_2, q_3)$.

Conceitos básicos sobre os quatérnios

- Note que $\mathbf{1}, \mathbf{i}, \mathbf{j}, \mathbf{k}$ formam uma base para o conjunto dos quatérnios, que é comumente denotado por $\mathbb{H}$.

- Um quatérnio também pode ser escrito na forma $q = q_0 + \vec{q}$, onde q_0 e $\vec{q}$ são chamados, respectivamente, parte real e parte vetorial de q.

- As partes real e vetorial de um quatérnio também podem ser denotadas por $Re\{q\} := q_0$ e $Ve\{q\} := \vec{q}$.

- A soma $p + q$ de dois quatérnios $p = p_0 + p_1\mathbf{i} + p_2\mathbf{j} + p_3\mathbf{k}$ e $q = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$ é o quatérnio obtido pela adição das suas componentes:

$$p + q = (p_0 + q_0) + (p_1 + q_1)\mathbf{i} + (p_2 + q_2)\mathbf{j} + (p_3 + q_3)\mathbf{k}$$

.

Conceitos básicos sobre os quatérnios

- O produto pq de dois quatérnios $p = p_0 + \vec{p}$ e $q = q_0 + \vec{q}$ é o quatérnio dado por:

$$pq = p_0 q_0 - \vec{p} \cdot \vec{q} + p_0 \vec{q} + q_0 \vec{p} + \vec{p} \times \vec{q}$$

  Note que o produto de dois quatérnios não é comutativo.

- O conjugado e a norma de um quatérnio q são definidos respectivamente por: $\bar{q} = q_0 - \vec{q}$ e $||q|| = \sqrt{\bar{q}q} = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}$

- Um quatérnio q é dito unitário quando $||q|| = 1$. Denotamos por $\mathbb{S}$ o conjunto de todos os quatérnios unitários, isto é:

$$\mathbb{S} = \{q \in \mathbb{H} : ||q|| = 1\}$$

- Perceba que $\mathbb{S}$ pode ser visto como uma hiperesfera unitária em $\mathbb{R}^4$.

Conceitos básicos sobre os quatérnios

- Para um número complexo $z = a + b\mathbf{i}$, o seu argumento principal é dado pela função $atan2(b, a)$. Temos $atan2(b, a) = arg(a + b\mathbf{i})$, onde $-\pi < atan2(b, a) \leq \pi$.

- Similarmente, para um quatérnio $q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$, temos três argumentos principais: $arg_{\mathbf{i}} = atan2(b, a)$, $arg_{\mathbf{j}} = atan2(c, a)$ e $arg_{\mathbf{k}} = atan2(d, a)$.

- Um quatérnio q pode ser alternativamente expresso usando a sua representação ângulo-fase:

$$q = ||q||e^{\mathbf{i}\phi}e^{\mathbf{k}\psi}e^{\mathbf{j}\theta}, \quad \phi \in [-\pi, \pi), \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}), \psi \in [-\frac{\pi}{4}, \frac{\pi}{4}) \qquad (4)$$

- Os números imaginários $\mathbf{i}$, $\mathbf{j}$ e $\mathbf{k}$ podem ser computados usando a fórmula de Euler: $e^{\mathbf{i}\phi} = \cos(\phi) + \mathbf{i}\mathrm{sen}(\phi)$, $e^{\mathbf{k}\psi} = \cos(\psi) + \mathbf{k}\mathrm{sen}(\psi)$ e $e^{\mathbf{j}\theta} = \cos(\theta) + \mathbf{j}\mathrm{sen}(\theta)$.

Uma representação ângulo-fase para um quatérnio

Teorema

Cada quatérnio q pode ser representado na forma

$$q = ||q||e^{\mathbf{i}\phi}e^{\mathbf{k}\psi}e^{\mathbf{j}\theta}, \quad \phi \in [-\pi, \pi), \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}), \psi \in [-\frac{\pi}{4}, \frac{\pi}{4}) \tag{5}$$

- Necessitamos de dois lemas para provar o teorema.

Lema 1

Seja $q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ *um quatérnio e* $\mathbb{M}$ *um mapeamento do conjunto dos quatérnios no conjunto das matrizes reais* 3×3 *dado por:*

$$\mathbb{M}(q) = \begin{bmatrix} a^2+b^2-c^2-d^2 & 2(bc-ad) & 2(bd+ac) \\ 2(bc+ad) & a^2-b^2+c^2-d^2 & 2(cd-ab) \\ 2(bd-ac) & 2(cd+ab) & a^2-b^2-c^2+d^2 \end{bmatrix}$$

Então: $\mathbb{M} : \{q \in \mathbb{H} : ||q|| = 1\} \to SO(3)$ *é um homomorfismo entre grupos SOBREJETIVO.*

Uma representação ângulo-fase para um quatérnio

Lema 2

Sejam q_1 e q_2 quatérnios unitários tais que $\mathbb{M}(q_1) = \mathbb{M}(q_2)$. Então, $q_1 = q_2$ ou $q_1 = -q_2$.

- Vamos agora demonstrar o teorema.
- Todo quatérnio q pode ser reescrito na forma $q = ||q||r$, onde r é um quatérnio unitário.
- Cada r possui uma matriz de rotação $R = \mathbb{M}(r)$ associada a ele.
- Podemos escrever $R = R_x(2\phi)R_z(2\psi)R_y(2\theta)$, onde R_x, R_z e R_y são as rotações em torno dos eixos coordenados dados em cada índice e 2ϕ, 2ψ e 2θ são denominados *ângulos de Eüler*.
- Sendo $q_1 = e^{\mathbf{i}\phi}$, $q_2 = e^{\mathbf{j}\theta}$ e $q_3 = e^{\mathbf{k}\psi}$ verifica-se que $R(q_1) = R_x(2\phi)$, $R(q_2) = R_y(2\theta)$ e $R(q_3) = R_z(2\psi)$.

Uma representação ângulo-fase para um quatérnio

- Como $\mathbb{M}$ é um homomorfismo entre grupos temos $\mathbb{M}(q_1)\mathbb{M}(q_3)\mathbb{M}(q_2) = \mathbb{M}(q_1 q_3 q_2) = \mathbb{M}(r)$.
- Do lema 2 temos que $r = q_1 q_3 q_2$ ou $r = -q_1 q_3 q_2$.
- Assim, obtemos $q = \pm ||q|| e^{\mathbf{i}\phi} e^{\mathbf{k}\psi} e^{\mathbf{j}\theta}$.
- Podemos eliminar o sinal "menos" trocando ϕ por $\phi + \pi$.
- Vamos agora determinar os intervalos mínimos de variação dos ângulos ϕ, ψ e θ de modo que possamos representar um quatérnio q, mas basta fazermos isto para r.

Uma representação ângulo-fase para um quatérnio

- Vamos construir a matriz de Rodriguez $R = \mathbb{M}(r)$, mas usando a relação $\mathbb{M}(r) = R_x(2\phi)R_z(2\psi)R_y(2\theta)$:

$$\begin{pmatrix} \cos(2\psi)\cos(2\theta) & -\text{sen}(2\psi) & \cos(2\psi)\text{sen}(2\theta) \\ \cos(2\phi)\text{sen}(2\psi)\cos(2\theta)+\text{sen}(2\psi)\text{sen}(2\theta) & \cos(2\phi)\cos(2\psi) & \cos(2\phi)\text{sen}(2\psi)\text{sen}(2\theta)-\text{sen}(2\psi)\cos(2\theta) \\ \text{sen}(2\phi)\text{sen}(2\psi)\cos(2\theta)-\cos(2\psi)\text{sen}(2\theta) & \text{sen}(2\phi)\cos(2\psi) & \text{sen}(2\phi)\text{sen}(2\psi)\text{sen}(2\theta)+\cos(2\psi)\cos(2\theta) \end{pmatrix}$$

- Analisando o elemento R_{12} da matriz obtemos:

$$-\text{sen}(2\psi) = 2(bc - ad) \Rightarrow \psi = -\frac{\text{arcsen}[2(bc - ad)]}{2}$$

- Portanto, $\psi \in [\frac{-\pi}{4}, \frac{\pi}{4}]$.

- As análises dos outros ângulos ocorrem de modo similar. Uma observação para essas análises é que elas devem ser feitas em três casos: $\psi \in]\frac{-\pi}{4}, \frac{\pi}{4}[$, $\psi = \frac{-\pi}{4}$ e $\psi = \frac{\pi}{4}$.

Rede de Hopfield com neurônios multiestados

- Cada vetor do conjunto de memórias fundamentais é um vetor com entradas quaterniônicas.
- O estado u_p de um neurônio da rede, com $||u_p|| = 1$ é dado por
$$u_p = e^{\mathbf{i}\phi_p} e^{\mathbf{k}\psi_p} e^{\mathbf{j}\theta_p}.$$
- Podemos reescrever u_p numa notação mais compacta fazendo $e^{\mathbf{i}\phi_p} = q^{(\phi)}$, $e^{\mathbf{k}\psi_p} = q^{(\psi)}$ e $e^{\mathbf{j}\theta_p} = q^{(\theta)}$:
$$u_p = q^{\phi_p} q^{\psi_p} q^{\theta_p}.$$
- O potencial de ativação de um neurônio p é dado por:
$$h_p(t) = \sum_q w_{pq} u_q(t) = \sum_q w_{pq} e^{\mathbf{i}\phi_q} e^{\mathbf{k}\psi_q} e^{\mathbf{j}\theta_q} = \sum_q w_{pq} q^{(\phi_q)}(t) q^{(\psi_q)}(t) q^{(\theta_q)}(t)$$

Rede de Hopfield com neurônios multiestados

- Como função de ativação neste modelo usamos a função sinal quaterniônica, representada por $qsign(u)$ e dada por:

$$qsign(u) = csign_A(q^{(\phi)})csign_B(q^{(\psi)})csign_C(q^{(\theta)})$$

- A função $csign_A(q^{(\phi)})$ é a imagem da função sinal-complexa aplicada ao número complexo $q^{(\phi)}$ e é dada por:

$$csign_A(q^{(\phi)}) = \begin{cases} -e^{0}, & -\pi \leq arg(q^{(\phi)}) < -\pi + \phi_0 \\ -e^{i\phi_0}, & -\pi + \phi_0 \leq arg(q^{(\phi)}) < -\pi + 2\phi_0 \\ \dots & \dots \\ -e^{i(A-1)\phi_0}, & -\pi + (A-1)\phi_0 \leq arg(q^{(\phi)}) < -\pi + A\phi_0, \end{cases} \tag{6}$$

onde $\phi_0 = \frac{2\pi}{A}$.

Rede de Hopfield com neurônios multiestados

- A função $csign_B(q^{(\psi)})$ é a imagem da função sinal-complexa aplicada ao número complexo $q^{(\psi)}$ e é dada por:

$$csign_B(q^{(\psi)}) = \begin{cases} e^{k\cdot(\frac{-\pi}{4})}, & \frac{-\pi}{4} \leq arg(q^{(\psi)}) < \frac{-\pi}{4} + \psi_0 \\ \dots & \dots \\ e^{k(\frac{-\pi}{4}+(B-1)\psi_0)}, & \frac{-\pi}{4} + (B-1)\psi_0 \leq arg(q^{(\psi)}) < \frac{-\pi}{4} + B\psi_0, \end{cases} \quad (7)$$

onde $\psi_0 = \frac{\pi}{2B}$.

- A função $csign_C(q^{(\theta)})$ é a imagem da função sinal-complexa aplicada ao número complexo $q^{(\theta)}$ e é dada por:

$$csign_C(q^{(\theta)}) = \begin{cases} e^{j\cdot(\frac{-\pi}{2})}, & \frac{-\pi}{2} \leq arg(q^{(\theta)}) < \frac{-\pi}{2} + \theta_0 \\ \dots & \dots \\ e^{j(\frac{-\pi}{2}+(C-1)\theta_0)}, & \frac{-\pi}{2} + (C-1)\theta_0 \leq arg(q^{(\theta)}) < \frac{-\pi}{2} + C\theta_0, \end{cases} \quad (8)$$

onde $\theta_0 = \frac{\pi}{C}$.

Rede de Hopfield com neurônios multiestados

- Postas essas definições, o estado do neurônio p no instante $t+1$ é dado por:

$$u_p(t+1) = qsign(h_p(t))$$

- Os neurônios devem ser atualizados de modo assíncrono com a condição que $q^{(\phi)}$, $q^{(\psi)}$, $q^{(\theta)}$ de um neurônio nunca sejam ativados simultaneamente.

- No caso de termos N neurônios, a função energia da rede é definida por:

$$E(t) = -\frac{1}{2}\sum_{p=1}^{N}\sum_{q=1}^{N} u_p^*(t) w_{pq} u_q(t),$$

com $w_{pq} = w_{qp}^*$ e $E = E^*$ e $w_{pp\geq 0}$.

Rede de Hopfield com neurônios multiestados

- Com a atualização apenas do r-ésimo neurônio da rede (e sem atualizar os três argumentos simultaneamente) obtemos:

$$u_p(t+1) = \begin{cases} q^{(\phi_p)}(t)q^{(\psi_p)}(t)q^{(\theta_p)}(t) = u_p(t), \quad p \neq r \\ \\ q^{(\phi_p)}(t)q^{(\psi_p)}(t+1)q^{(\theta_p)}(t+1) \\ \qquad\qquad ou \\ q^{(\phi_p)}(t+1)q^{(\psi_p)}(t+1)q^{(\theta_p)}(t), \quad p = r. \end{cases}$$