Proceedings of the V Brazilian Conference on Neural Networks - V Congresso Brasileiro de Redes Neurais
pp. 235–240, April 2–5, 2001 - Rio de Janeiro - RJ - Brazil

Um Controlador da Dinâmica da Aritmética de Múltiplas Colunas usando Redes Neurais Artificiais

Andréa M. L. Doescher[1], Orlando Bisacchi Coelho[2], Sandra Sandri[3]
[1]LAC/INPE, Cx Postal 515, 12201-970 S.J.Campos /SP- Brazil
[2] Universidade de Mogi das Cruzes (UMC)
[3]LAC/INPE, Cx Postal 515, 12201-970 S.J.Campos /SP- Brazil
E-mails: andreald@bol.com.br, orlando@umc.br, sandri@lac.inpe.br

Abstract

This work describes a series of improvements on Dallaway´s neural network-based model for multiple column arithmetics. Both the original model and the improvements herein reported are firmly rooted in data collected in studies on how children are actually taught and learn arithmetics. Several changes in the training data, the training regime and the actions that control the execution of the arithmetic operations – which are the neural network´s outputs – were carried on. This way a more adequate model was developed, in terms of the generalization ability of the network that underlies the model, the correction of the results produced and, most important, the cognitive plausibility of the errors generated by the model.

1. Introdução

A cognição humana tem sido investigada há mais de 2000 anos, mas apenas nos últimos 100 anos pode-se dizer que ela tem sido estudada cientificamente. O conhecimento dos processos cognitivos humanos aumentou significativamente nos últimos 45 anos devido, em parte, ao avanço tecnológico dos computadores digitais. Dentre esses processos cognitivos, a aritmética tem sido modelada utilizando-se técnicas computacionais como sistemas de produção e redes neurais artificiais [1].

As redes neurais, em parte por serem modelos computacionais inspirados no sistema nervoso central humano, apresentam características próximas às encontradas em processos cognitivos humanos (por exemplo: aprendem pela experiência, generalizam a partir de exemplos e abstraem conceitos), características estas que não são exibidas por um sistema de produção. Essas características demonstram a relevância do uso de redes neurais na modelagem dos processos cognitivos humanos [9]. No entanto, essa tarefa não é simples. Em se tratando da modelagem aritmética de múltiplas colunas, essa tarefa é particularmente complexa [4, 5, 6,9,13]. Isto porque:

- A aritmética de múltiplas colunas envolve processamento seqüencial. Assim, para que uma rede neural modele este tipo de processamento é necessário que a rede seja capaz de processar seqüências.
- Além da necessidade de processar seqüências, a rede neural também deve aprender a regra implícita contida em uma operação aritmética de múltiplas colunas, uma vez que é impossível (e seria mesmo indesejável) treinar a rede para todos os de casos possíveis que uma operação aritmética de múltiplas colunas possa gerar.

As longas multiplicações raramente são efetuadas mentalmente pelo homem; a grande maioria utiliza uma representação externa no papel. A multiplicação de múltiplas colunas, sobretudo, é feita utilizando-se uma representação externa, onde leva-se em consideração a forma com que dispomos algarismos e símbolos no papel e a capacidade de nosso olhos se moverem e focarem nossa atenção em diferentes áreas do papel [11]. Aprender a multiplicação de múltiplas colunas é essencialmente aprender a controlar a execução dessas ações apropriadas na seqüência correta.

O modelo de Dallaway usa redes neurais artificiais para modelar as ações visuais e mentais que são realizadas ao se efetuar uma operação aritmética de múltiplas colunas no papel. No presente trabalho, nós modelamos o comportamento humano na execução da aritmética de múltiplas colunas (especificamente a adição e a multiplicação), a partir do modelo de Dallaway. A abordagem proposta aqui sana algumas das deficiências apresentadas pelo modelo de Dallaway. Além disso transportamos para a rede técnicas pedagógicas utilizadas para ensinar aritmética à criança brasileira [2], visando aproximar ainda mais o comportamento da rede do comportamento humano.

O trabalho está organizado da seguinte maneira. A arquitetura do modelo de Dallaway está descrita na próxima seção. Na seção 3 são descritas algumas das simulações utilizando a nova abordagem, e a seção 4 traz a conclusão.

2. Modelo Básico

O presente trabalho utiliza como base o modelo de Dallaway [5]. Neste modelo, a representação de um problema de aritmética é feita por meio de uma grade, onde a rede utiliza um foco de atenção para atender às áreas particulares do problema (ler ou escrever um dígito) ao longo da grade. Além do Foco de Atenção (FA), são usados três registradores externos que indicam o progresso da rede através do problema: Registrador Superior (RS), Registrador Inferior (RI) e Registrador da Resposta (RR). Há um outro registrador, chamado Registro do Foco de Atenção (RFA), cuja célula pode ser armazenada e usada em cálculos. Quando a operação de multiplicação ou adição é chamada, o cálculo usa a célula em foco (FA) e a última célula armazenada no RFA, colocando o resultado em um acumulador (ACM).

Há um conjunto de vinte e quatro ações (operações) de saída que podem ser executadas tanto na multiplicação como na adição, como por exemplo: adicionar valor do FA ao acumulador (AFA); calcular Produto (CPR); escrever unidades (EUN); fim da operação (FTO); marcar dígito vai-um (MVU) entre outras. Para cada operação há uma unidade de saída associada à rede. A rede não efetua as operações (ações) necessárias para que a operação aritmética seja realizada, mas age como um controlador dessas operações. Estas ações dirigem mecanismos externos para atualizar o problema de várias formas, como escrever um número ou calcular produto.

A figura 1 traz um exemplo da representação e da dinâmica de procedimentos para o problema 2x3, e a tabela 1 traz o conjunto de ações respectivo.

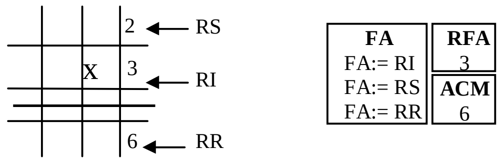

Fig. 1. Exemplo da representação do problema 2x3.

Tabela 1. Conjunto de ações de saída do problema 2x3

Ações	ários
IMU	Os registradores se posicionam e o número 3 é ção, isto é, FA apontará número que o RI está apontando
ARM	O número 3 é colocado no RFA
PRS	Move o FA para a posição indicada por RS
CPR	Multiplica os dígitos que estão no RFA com os que ão no FA e coloque a resposta no ACM
PPR	Move o FA para a posição indicada por RR
EUN	Escreve a resposta do ACM na posição indicada por FA
FTO	Terminou a operação.

A arquitetura do modelo (Figura 2) consiste de uma rede recorrente de três camadas (com 9, 35 e 24 neurônios nas camadas de entrada, escondida e de saída, respectivamente), treinada com o algoritmo BPTT (backpropagation through time). A camada escondida é copiada para a camada de contexto (com 35 neurônios), a cada passo, agindo como uma memória do estado anterior.

Os dados de entrada da rede (Figura 3) são representados por um vetor de 9 bits; a tarefa dos bits é somente de estabelecer o início do processamento; indicando o que a célula corrente contém, que pode ser: um dígito, um espaço em branco ou um traço.

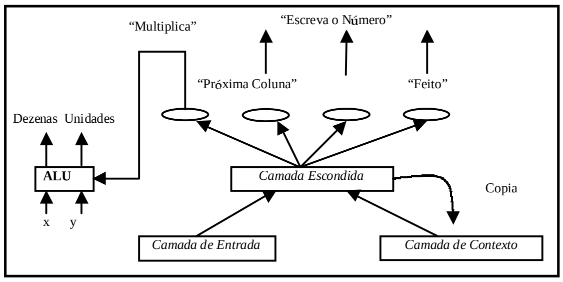

Fig. 2. Arquitetura do modelo / Fonte: Dallaway [5]

O conjunto de treinamento contém apenas um único problema de cada instância dos problemas aritméticos, devido a todos produzirem exatamente a mesma seqüência de vetores de saída. Esses problemas foram escolhidos e organizados de acordo com o grau de dificuldade encontrado em livros escolares (vide Tabela 2).

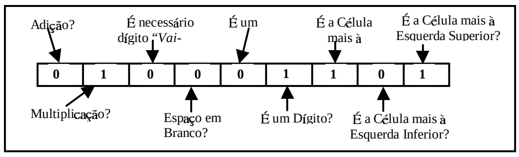

Fig. 3. Representação dos dados de entrada [5]

Tabela 2. Problemas utilizados para treinar a rede

1: 1+1	**8:** 101+109	**15:** 12x5	**22:** 12x50
2: 1+1+1	**9**: 101+99	**16:** 12x9	**23:** 12x55
3:11+11	**10:** 101+899	**17:** 1x11	**24:** 12x59
4: 11+1	**11:** 1x1	**18:** 11x11	**25:** 12x90
5: 1+9	**12:** 2x5	**19:** 1x111	**26:** 12x95
6: 1+19	**13:** 11x1	**20:** 12x15	**27:** 12x99
7: 100+100	**14:** 111x1	**21:** 12x19	**28:** 111x11

Adaptada de Dallaway [5]

O treinamento da rede é feito incrementalmente: treina-se a rede para o 1º problema e após ser treinada acrescenta-se o 2º problema ao 1º e treina-se a rede para esse novo conjunto de treinamento (1º + 2º problema); esse processo se repete até que todos os problemas de treinamento sejam incorporados ao conjunto de treinamento inicial. Treina-se a rede até que uma das seguintes condições seja satisfeita:

a) $max_i \, |y_i - d_i| \leq 0{,}2$ onde y_i é a saída obtida e d_i a saída desejada da rede

b) são atingidas 60.000 épocas de treinamento. Neste caso, é introduzido um novo conjunto de treinamento e o aprendizado continua.

É feita uma análise dos erros cometidos pela rede (nos testes de generalização). Os erros estão divididos em dois grupos: erros observados e não observados. Os *Erros Observados* (EO) são aqueles erros que foram encontrados na literatura. Já os *Erros Não-Observados*

(ENO) são aqueles erros que não são diagnosticados em humanos, isto é, não são encontrados na literatura. Os erros não observados estão subdivididos em *Erro de Máquina* (EM), que uma criança jamais cometeria e *Erro Plausível* (EP), erro que não foi catalogado, mas que uma criança poderia cometer.

3. Nova Abordagem

Consideramos como *comportamento semelhante ao do homem* a soma da freqüência de generalizações corretas mais a de EO que foram catalogados por: Buswell, Cox, Ainsworth e Attisha, e foram listados em Dallaway [5].

Buswel testou 263 crianças do terceiro ao sexto ano do ensino fundamental; Cox testou 564 crianças do segundo ao sexto ano do ensino fundamental; Ainsworth trabalhou apenas com a multiplicação, testando 76 crianças que foram divididas em dois grupos: um de 8 a 9 anos e o outro com crianças de 10 a 11 anos; Attisha não diz a freqüência da ocorrência de cada erro, embora liste um grande número de erros [5].

O desempenho do modelo de Dallaway para aritmética de múltiplas colunas é considerado baixo, pois a rede ocasiona muitos erros quando testada em problemas não vistos anteriormente, sendo que em 51,43% dos casos o comportamento é semelhante ao do homem (14,29% de generalizações corretas + 37,14% de EO) e que em 48,59% do comportamento é de máquina [5]. Quanto ao comportamento global do modelo, ele também é considerado baixo, pois a rede, após ser treinada em todos os problemas do conjunto de treinamento, não consegue generalizar problemas com quatro dígitos no multiplicando em nenhum momento, além de não efetuar corretamente problemas cujo multiplicador tenha três ou mais dígitos.

Tendo como meta reproduzir o mais fielmente possível o comportamento humano na resolução de um problema aritmético, foram feitas várias alterações no modelo de Dallaway (descritas nos próximos parágrafos). Além disso, foram usadas técnicas pedagógicas do ensino aritmético no treinamento da rede [6].

Sendo o treinamento incremental, ao passar de um problema de treinamento para outro, todos os pesos de todas as camadas ficam ativos e a rede inicia seu treinamento utilizando esses pesos. Contudo, cada problema do conjunto de treinamento, ao ser executado separadamente, recebe o valor zero dos nós da camada de contexto. Sendo assim, a rede se perde na execução do problema (pois ela é treinada de uma forma e executada de outra). Para resolver esse problema, a camada de contexto é zerada sempre que um novo problema for incorporado ao treinamento da rede, em todos os experimentos realizados.

Foram estabelecidos padrões para gerar as ações de saída de cada problema do conjunto de treinamento. Além disso, a generalização da rede foi testada para os três problemas consecutivos (da mesma operação) de cada problema do conjunto de treinamento.

No primeiro experimento realizado, a rede foi treinada utilizando uma taxa de aprendizado de 0,01 para os problemas de 1 a 24 e uma taxa de 0,001 para os problemas restantes. A rede generalizou 13 dos 72 problemas. O comportamento da rede é bom, pois em 84,73% dos casos o comportamento é semelhante ao do homem (em Dallaway o valor é de 51,43%) e em 15,28% é de máquina.

A partir do comportamento do modelo, foi observado que, por haver dois padrões distintos para a ocorrência de um dígito vai-um (um para problemas de uma coluna como 1+9 e 2x5 e um outro padrão para problemas de duas ou mais colunas como 1+19, 12x5, 12x19 e outros) e na leitura das células, a rede se perdia e cometia erros. Para resolver esse problema, foi adotado um único padrão tanto na ocorrência de um dígito vai-um (independentemente do número de colunas do problema que originou o dígito vai-um) como para leitura da célula (independentemente dela conter um dígito ou ser uma célula em branco).

Além disso, analisando livros escolares e as propostas de Atividades Matemáticas [2], foi verificado que as crianças aprendem a somar 11+1, 12+1, 13+1 etc., antes de somar problemas como 11+11. Desta forma, foi alterado a ordem dos problemas 3 e 4 (ver Tabela 1). Com todas essas mudanças, o comportamento da rede melhorou sensivelmente, de tal maneira que 29 dentre 73 problemas foram generalizados. Além disso, o comportamento do modelo é semelhante ao do homem em 80,83% dos casos (39,73% correto + 41,1% EO) e de máquina em 19,18%. Embora o comportamento do modelo em relação ao comportamento humano tenha diminuído em aproximadamente 4%, houve um grande aumento (21,67%) em relação ao número de problemas que a rede generalizou corretamente.

É sabido que qualquer multiplicação de dois números naturais pode ser transformada em uma soma. Entretanto, dependendo do multiplicador, o número de fatores desta soma pode ser elevado. Por exemplo, a multiplicação de 456x115 é equivalente a uma soma contendo 115 fatores iguais a 456. Embora factível, este processo não é nada prático. Um outro algoritmo, ou técnica operatória, transforma a multiplicação em uma soma contendo um número de fatores igual ao número de dígitos do multiplicador. Para o exemplo acima temos: 456 x 115 = 2280 + 4560 + 45600 = 52440 [2]. É observado que este algoritmo pode ser subdividido em duas etapas: geração dos fatores para a adição e a adição desses fatores. Esta segunda etapa (adição dos fatores) coincide com a adição como operação isolada. Isto demonstra que, no processo da multiplicação, a única etapa que realmente necessita ser aprendida é a geração dos fatores, pois a adição foi ensinada anteriormente. Portanto, a rede necessita aprender que, quando os fatores já foram gerados, ela não mais necessita das informações a respeito da geração dos fatores e para não propagar essas informações adiante, causando erros na adição, é necessário zerar a camada de contexto sempre que for ser iniciada a adição das parcelas do produto. Depois dessa alteração, a rede generalizou 30 problemas no total de 73 e o número de erros não observados baixou de 19,18% para 12,33%. O comportamento da rede é semelhante ao do homem em 87,68% dos casos (41,1% correto + 46,58% EO) e de máquina em 12,33%.

Ao se zerar a camada de contexto no início de uma soma decorrente de uma multiplicação, cada conjunto de treinamento da multiplicação (cujo multiplicador tem duas ou mais colunas) foi transformado em dois problemas de treinamento agrupados, um de multiplicação e outro da adição envolvida. Se a rede for adequadamente treinada na adição, não há necessidade de se repetir este treinamento nas multiplicações. Desta forma, foram retiradas as ações referentes às somas dos problemas de treinamento das multiplicações e o resultado foi que a rede generalizou 30 problemas no total de 73 (como no experimento anterior) e a freqüência dos erros de máquina baixou de 12,33% para 1,37%. O comportamento da rede é semelhante ao do homem em 97,26% dos casos (41,1% correto e 56,16% EO) e de máquina 2,74%.

Embora a retirada das ações referentes às adições dos problemas de multiplicação não tenha trazido nenhum prejuízo para a rede em relação ao aprendizado e à generalização, a rede erra as somas das parcelas do produto (no teste de generalização) de vários problemas, o que nos sugere que o número de problemas de treinamento da adição não é suficiente para que a rede aprenda e realize uma multiplicação qualquer. Além disso, na generalização de alguns problemas, a rede se perde ao passar o traço no final de uma multiplicação, somando incorretamente as parcelas do produto.

Uma criança observando atentamente as parcelas do produto, sabe qual é o último dígito a ser adicionado. No caso da rede, esta informação é passada pelo traço da multiplicação e quando ele não cobre todas as colunas da parcela do produto, a rede erra a soma dessas parcelas, pois finaliza a operação de acordo com a informação contida no traço. Para resolver este problema, foi feita uma modificação na forma em que a rede passa o traço depois de uma multiplicação, para garantir que todas as colunas contendo dígitos sejam cobertas pelo traço. Com essa modificação a rede generalizou pela primeira vez o problema 111x11, perfazendo um total de 31 problemas e os erros cometidos nos testes de generalizações não são de máquina - o que por si só é um bom resultado. Além disso, 98,63% dos casos (42,47% correto e 56,16% EO) do comportamento da rede é semelhante ao do homem e 1,37% é de máquina (especificamente EP).

Como foi visto no experimento anterior que os problemas de adição do conjunto de treinamento não eram suficientes para delinear cada instância de problemas aritméticos. Por esta razão, inserimos novos problemas ao conjunto de treinamento (Tabela 3).

Tabela 3. Novo conjunto de treinamento

1: 1+1	**14:**999+9999+9999	**27:** 12x15
2: 1+1+1	**15:** 1x1	**28:** 12x19
3:11+1	**16:** 2x5	**29:** 12x50
4: 11+11	**17:** 11x1	**30:** 12x55
5: 1+9	**18:** 111x1	**31:** 12x59
6: 19+1	**19:** 12x5	**32:** 12x90
7: 100+1	**20:** 12x9	**33:** 12x95
8: 101+109	**21:** 1x11	**34:** 12x99
9: 101+99	**22:** 11x11	**35:** 99x99
10: 101+899	**23:** 111x11	**36:** 199x99
11: 9999+1	**24:** 1x111	**37:** 99x999
12: 9999+9999	**25:** 11x111	**38:** 199x999
13: 1+10+100	**26:**111x111	

Com a inserção desses novos problemas, a rede generalizou 54 problemas dentre os 119 problemas testes. O número de problemas generalizados corretamente (45,38%) é o maior até aqui. Essa melhora não é apenas em relação aos problemas generalizados (a rede generaliza problemas com mais de quatro dígitos no multiplicando), mas também em relação ao tipo de erro cometido pela rede. No primeiro experimento a taxa de erros de máquina é de 12,5%, enquanto que agora essa taxa é de 6,72%. Em suma, todas as mudanças efetuadas até se chegar a esse experimento colaboraram tanto para que a rede tivesse um melhor desempenho (do ponto de vista computacional) como também para que seu comportamento fosse mais próximo ao do homem.

Apesar de todas as modificações até aqui, em nenhum momento a rede consegue generalizar problemas cujo multiplicador tenha quatro ou mais dígitos. Através dos diversos testes feitos chegou-se a conclusão que, para resolver este problema, tem-se as seguintes opções:

1) É necessário que a rede seja treinada neste tipo de problema: para isso basta acrescentar ao conjunto de treinamento problemas com o número de dígitos no multiplicador que se queira que ela execute. No entanto, dependendo do problema que se queira que a rede generalize, o conjunto de treinamento pode ficar muito grande;

2) A rede precisa aprender a regra intrínseca aos zeros que são colocados na linha de resposta, de acordo com o multiplicador: baseando no modelo humano, essa opção é a mais indicada. No entanto, apesar de diversos testes, não chegou-se a bons resultados.

3) Utilizar um artifício computacional que indique à rede quantos zeros e em quais linhas de resposta eles devem ser colocados: essa opção resolve o problema, mas sem considerar as questões cognitivas.

4. Conclusão

A utilização de técnicas pedagógicas do ensino da aritmética na modelagem da aritmética por redes neurais aproximaram sensivelmente o comportamento do modelo do comportamento humano. Com efeito, na maioria dos testes realizados a rede acertou aonde o homem geralmente acerta e errou também, aonde o homem geralmente erra. Isto nos leva a crer que as redes neurais podem ser utilizadas não apenas para modelar processos cognitivos humanos, mas que podem ser utilizadas como plataformas para testar novas técnicas pedagógicas do ensino matemático.

Na tentativa de se modelar o comportamento humano na resolução de problemas aritméticos, vários experimentos foram realizados e, através deles, foi observado que vários são os fatores que contribuem para que uma rede neural artificial possa modelar a aritmética de múltiplas colunas, tal qual efetuada pelo homem:

- A adoção dos padrões de operações utilizados para gerar os problemas de treinamento pode interferir no desempenho da rede;
- É importante que o conjunto de treinamento utilizado instancie corretamente os problemas aritméticos;
- A ordem dos problemas do conjunto de treinamento interfere na generalização e comportamento da rede;
- A única etapa que realmente necessita ser aprendida na multiplicação é a geração dos fatores;
- A forma de se passar o traço da multiplicação é importante para que a rede generalize corretamente;
- Em termos cognitivos, é difícil que a rede generalize problemas com mais de 3 dígitos no multiplicador.

Considerando que o comportamento do modelo em 84,03% dos casos (no pior resultado) é semelhante ao do homem, podemos concordar com o que diz Anderson [1], no intuito que outros conceitos matemáticos sejam modelados, utilizando o escopo deste trabalho:

"...embora as arquiteturas das redes sejam diferentes, todas elas apresentaram freqüentes erros e estes erros não foram aleatórios – os erros foram próximos aos erros ocasionados pelos humanos. Além disso, cada um dos modelos mostrou razoáveis adequações a várias descobertas experimentais bem estabelecidas para a aritmética, o que indica que o paradigma de redes neurais é apropriado para a modelagem da recuperação dos fatos aritméticos" (p. 570).

Referências Bibliográficas

[1] Anderson, J.A, *Mental Arithmetic Using Neural Networks.* The Handbook of Brain Theory and Neural Networks, Arbib, M. A. (ed.), The MIT Press, 1995, pp. 570-572.

[2] *Atividades Matemáticas*, Secretaria de Estado da Educação - São Paulo, Coordenadoria de Estudos e Normas Pedagógicas. Governo do Estado de São Paulo, vol. 1, 3ª série do 1º Grau & 4ª série do 1º Grau, 1991.

[3] Cotofana, S. e Vassiliadis, S. Serial binary multiplication with feed-forward neural networks. *Neurocomputing*, v. 28, p. 1-19, october, 1999.

[4] Cottrell, G.W & Tsung, F., *Learning Simple Arithmetic Procedures*, in J.A. Barnden & J. B. Pollack (eds.), Advances in Connectionist and Neural Computation Theory, Vol. 1, High-Level Connectionist Models, Cap. 12, pp. 305-321, Norwood, Ablex, 1991.

[5] Dallaway, R., *Dynamics of Arithmetic: A Connectionist View of Arithmetic Skills.* Cognitive Science Research Papers no. 306, University of Sussex at Brighton, Brighton, 1994.

[6] Doescher, A.M.L. Um Controlador da Dinâmica da Aritmética de Múltiplas Colunas usando Redes Neurais Artificiais. Dissertação de Mestrado, INPE, São José dos Campos, 2000.

[7] Hecht, S.A. Toward and informatio-processing account of individual differences in fraction skills. *Journal of Education Psychology*, V. 90, n. 3, p. 545-559, september, 1998.

[8] Mastropieri, M.A., Scruggs, T.E. e Shiah, R.L. Can computers teach problem-solving strategies to students with mild mental retardation? A case study. *Remedial and Special Education*, v. 18, n. 3, p. 157-165, may-jun, 1997.

[9] McCloskey, M. & Lindermann, M.A., *MATHNET: Preliminary Results from a Distributed Model of Arithmetic Fact Retrieval.*, in J. I. D. Campbell (ed.), The Nature and Origins of Mathematical Skills, Elsevier Science Publ., pp. 365-409, 1992.

[10] Recio, T. e Gonzalez-Lopes, M.J. Does computer algebra help at all learning real numbers? *Mathematics and Computer in Simulation*, v. 45, p.185-195, january, 1998.

[11] Suppes, P., *EyeMovement Models for Arithmetic and Reading Performance*, in E. Kowler (ed.), Eye Movements and their Role in Visual and Cognitive Processes, Cap. 10, pp. 455-477, 1990.

[12] VanLehn, K., Rule Acquisition Events in the Discovery of Problem-Solving Strategies. *Cognitive Science*, Vol. 15, pp. 1-47, 1991.

[13] Viscuso, S.R., Anderson, J., Rossen, M.L. & Sereno, M.E., *Experiments with Representation in Neural Networks: object, motion, speech, and arithmetic*, in J. A. Anderson, A. Pellionisz & E. Rosenfeld (eds.), Neurocomputing 2 – Directions For Research, pp. 700-716, 1990.

[15] Young, R.M. & O'Shea, T., Errors in Children's Subtraction, *Cognitive Science*, Vol. 5, pp. 153-177, 1981.