

UNIVERSIDADE FEDERAL DO PARANÁ
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
CURSO DE ENGENHARIA ELÉTRICA

FELIPE SILVEIRA GRECO

**REDES NEURAIS PARA RECONHECIMENTO DE
PADRÕES DE VÍDEO MPEG4**

CURITIBA

2011

FELIPE SILVEIRA GRECO

**REDES NEURAIS PARA RECONHECIMENTO DE
PADRÕES DE VÍDEO MPEG4**

Trabalho de Conclusão de Curso de
Engenharia Elétrica, Departamento de
Engenharia Elétrica, Setor de Tecnologia,
Universidade Federal do Paraná.

Orientador: Prof. Dr. Carlos Marcelo Pedroso

CURITIBA

2011

FELIPE SILVEIRA GRECO

**REDES NEURAIS PARA RECONHECIMENTO DE
PADRÕES DE VÍDEO MPEG4**

TRABALHO APRESENTADO AO CURSO DE ENGENHARIA ELÉTRICA, DA
UNIVERSIDADE FEDERAL DO PARANÁ, COMO REQUISITO PARCIAL À
OBTENÇÃO DO TÍTULO DE GRADUAÇÃO.

COMISSÃO EXAMINADORA

Prof. Dr. Carlos Marcelo Pedroso

Prof. Dr. Eduardo Parente Ribeiro

Prof. Eng. Roberto Heinrich

CURITIBA, JUNHO DE 2011.

RESUMO

O IPTV surge inserido no contexto de convergência de serviços no mundo das telecomunicações, onde se busca reunir todo tipo de mídia num único meio físico: a rede IP. Porém os desafios de realizar a tarefa de transmitir conteúdos de vídeo na rede da internet, em tempo real, com mesma qualidade que os serviços de televisão por assinatura oferecem hoje, são grandes. Redes Neurais são utilizadas como uma ferramenta para reconhecimento de padrões, partindo de princípios de funcionamento análogos ao do cérebro. Esta monografia tem como objetivo utilizar esta ferramenta para um estudo aplicado ao contexto do IPTV, procurando o melhor tipo de rede que reconheça padrões de tráfego de vídeo MPEG4. A proposta é de os resultados sejam utilizados como base para encontrar um algoritmo capaz de realizar descartes seletivos em roteadores e que seja mais eficaz que os outros algoritmos já utilizados hoje.

Palavras chave: IPTV, Redes Neurais, MPEG4.

ABSTRACT

The IPTV appears inserted in the context of convergence of services in the world of the telecommunications, where it searches to congregate all type of media in a single environment: IP network. However the challenges to perform such task (to transmit contents of video on the IP network ,in real time, with same quality that the services of television for signature offer today, are considerable. Neural networks are used as a tool for recognition of standards, considering analogous principles of brain functions. This monograph has as objective to use this tool for a study applied to the context of the IPTV, looking for optimum type of network that recognizes standards of MPEG4 video traffic. The proposal is that the results could be used as base to find an algorithm capable to carry through selective discardings in routers more efficiently than the others algorithms used nowadays.

Key words: IPTV, Neural Nets, MPEG4.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	CENÁRIO IPTV	1
1.2	OBJETIVOS E JUSTIFICATIVA	2
1.3	MOTIVAÇÕES	3
1.4	ESTRUTURA DA MONOGRAFIA	4
2	O SISTEMA IPTV	5
2.1	ARQUITETURA DE REFERÊNCIA IPTV	5
2.2	HEADEND	6
2.2.1	COMPRESSÃO DE VÍDEO MPEG4 (MOVING PICTURE EXPERT GROUP)	7
2.3	NÚCLEO DA REDE	9
2.3.1	PROTOCOLOS DE TRANSPORTE	9
	Protocolo UDP	10
	Protocolo de controle RTP	10
3	MECANISMOS DE GERÊNCIA DE FILAS	12
3.1	FIFO	12
3.2	RED	13
4	REDES NEURAIS ARTIFICIAIS	16
4.1	PRINCÍPIOS DE FUNCIONAMENTO	16
4.2	FUNÇÕES DE ATIVAÇÃO	17
4.3	TOPOLOGIAS DE REDE	18
4.4	APRENDIZADO	20
4.4.1	FUNDAMENTOS DO ALGORITMO <i>BACKPROPAGATION</i>	21
5	METODOLOGIA E RESULTADOS	25
5.1	METODOLOGIA	25
5.2	RESULTADOS	27
6	CONCLUSÕES	29
	REFERÊNCIAS BIBLIOGRÁFICAS	30

1 INTRODUÇÃO

1.1 CENÁRIO IPTV

É notável o crescimento da demanda por soluções convergentes quando tratamos de conteúdo multimídia transmitidos em tempo real pela Internet. Quando fala-se em convergência de serviços trata-se da consolidação do uso redes IP para todos os tipos de serviços, como telefonia e TV, além de dados.

Foi em meados dos anos 90 que a Internet passou a fazer parte do cotidiano de usuários domésticos. A partir daí, um novo cenário começou a ser traçado como desenvolvimento de novas tecnologias de acesso e transmissão, que proporcionaram o incremento na velocidade de conexão residencial e incentivaram o aumento do uso de aplicações multimídia nas redes IP (convergência). Um dos próximos passos em convergência seria o uso da rede IP para transmissão de sinais televisivos, o IPTV (*Internet Protocol Television*).

A motivação deste crescimento na demanda por serviços multimídia em ambiente doméstico se dá tanto por fatores econômicos e tecnológicos, como culturais.

Econômico porque o custo de implementação e manutenção de redes distintas para voz, dados e vídeo passou a não fazer sentido diante deste novo cenário. Cultural porque os usuários estão cada vez mais tendenciosos a exigirem a integração entre estas mídias, valorizando a oferta de serviços *triple play* pelas operadoras.

Já no aspecto técnico é possível citar as várias tecnologias de acesso e transmissão - xDSL (*Digital Subscriber Line*), WDM (*Wavelength-division Multiplexing*), FTTB (*Fiber to the Building*), FTTH (*Fiber to the Home*), etc - como fatores motivadores à convergência de serviços, pois estas tecnologias permitiram disponibilizar acesso de banda larga à rede, tanto nos *backbones* de longa distância, como em enlaces próximos ao usuário final.

Não podemos deixar de lado a influência do avanço nas tecnologias de compressão de áudio e vídeo no crescimento da utilização de serviços multimídia. Um exemplo bastante atual de padrão de compressão é o H.264/MPEG-4 AVC

(AVC – *Advanced Video Coding*), que tem se tornado o padrão de codificação de vídeo mais aceito desde a implantação do MPEG2 [1, 2, 3].

O IPTV está inserido neste contexto e pode ser definido como um serviço que possibilita a transmissão de conteúdo multimídia (vídeo, áudio e dados) sobre uma rede IP com garantias de qualidade, segurança, integridade e interatividade. O conteúdo é comprimido, encapsulado em protocolo IP e transmitido ao usuário em tempo real utilizando-se dos protocolos de transporte UDP/RTP.

O serviço IPTV é sensível a perdas e atrasos de pacotes, pois isto tem conseqüências na percepção de qualidade pelo usuário ou experiência do usuário (QoE - *Quality of Experience*). As operadoras de telecomunicações devem atentar muito a isto de forma a tornar o referido serviço competitivo diante da TV a cabo e da TV convencional.

Melhorar mecanismos de qualidade de serviços (QoS) reflete nos níveis de QoE. Assim, deve-se ter em vista a arquitetura geral do sistema de forma que se possa aperfeiçoar o desempenho de pontos estratégicos da rede para que possamos atingir qualidade equivalente aos serviços convencionais oferecidos hoje, ou mesmo níveis de qualidade mais elevados.

Um destes pontos seria evitar ao máximo a perda de pacotes que transportam conteúdo IPTV pela rede, evitando descontinuidade de sequência de imagens, que é a característica de maior impacto na qualidade de vídeo observada pelo usuário.

1.2 OBJETIVOS E JUSTIFICATIVA

Este projeto está situado dentro de um projeto de mestrado, onde está sendo proposto um algoritmo de gerência de filas, para ser implementado em roteadores que transportam filmes codificados com MPEG2 e MPEG4. Espera-se desenvolver um algoritmo para descarte seletivo de pacotes, que comparado aos algoritmos existentes, melhoraria a percepção de qualidade do usuário.

O objetivo específico desta monografia de fim de curso é testar diversas topologias de redes neurais para a identificação da carga de pacotes, de acordo com o tipo de quadro (P, I e B) dentro de um fluxo de transmissão de vídeo codificado com o *codec* MPEG4.

Evitar a perda de pacotes pertencentes a quadros do tipo I (conforme será explicado numa próxima sessão) durante o processo de roteamento pode melhorar qualitativamente a percepção da imagem por parte do usuário.

Será utilizado um conjunto de dados conhecido para que se possa treinar a Rede Neural para reconhecer pacotes utilizando a série temporal representada pelo intervalo entre pacotes e seus tamanhos. A Rede Neural deve identificar a presença de pacotes pertencentes aos quadros I do fluxo de quadros MPEG4.

1.3 MOTIVAÇÕES

Primeiramente deve-se pensar no motivo de tanto interesse das grandes operadoras de telecomunicações em atuar no ramo de IPTV. A resposta vem do fato já comentado anteriormente neste texto. O desenvolvimento tecnológico permite altas taxas de conexão na residência dos usuários, o que possibilitou a diversificação dos serviços pelas operadoras.

Além disto, a possibilidade de se fazer chamadas de voz via Internet (VoIP), entre outras aplicações, tende a diminuir a receita dos provedores de serviços de telefonia fixa neste ramo. Diversificar é preciso, e possível.

A penetração do IPTV no mundo cresce, sendo que no segundo semestre de 2007 a Europa já atingia mais de 4 milhões de assinantes, enquanto os assinantes na China cresceu 20,3% atingindo a marca de 736mil usuários [4]. No Brasil, as operadoras de telecomunicações ainda não podem disponibilizar conteúdo de televisão via rede IP por questões regulatórias.

O assinante de TV por assinatura está habituado com certo nível de qualidade percebida e não admite que um novo serviço ofereça qualidade inferior. Portanto, deve ser possível que o conteúdo de vídeo disponibilizado através da Internet chegue ao ponto de acesso do assinante com imagens de alta definição, assim como é oferecido nas transmissões via satélite ou cabo.

A temática principal deste trabalho é justamente possibilitar uma melhora qualitativa na entrega do conteúdo IPTV ao usuário final, de modo a melhorar a experiência deste e impulsionar a viabilidade deste modelo de negócio.

1.4 ESTRUTURA DA MONOGRAFIA

Esta monografia está dividida da seguinte forma: o primeiro capítulo apresenta uma visão geral sobre IPTV, dando uma visão geral do cenário do modelo de negócios atualmente.

O segundo capítulo referencia o funcionamento do sistema IPTV de forma geral, com objetivo de situar o leitor na problemática e apresenta os elementos básicos para o seu funcionamento. Além disto, este capítulo apresenta conceitos básicos do algoritmo de compressão de vídeo (MPEG4).

O terceiro capítulo é uma sessão reservada a tratar dos algoritmos de gerência de filas dos roteadores, mostrando de forma geral como são gerenciados os pacotes na rede IP e evidenciando os algoritmos FIFO e RED.

A seguir, no capítulo quatro, fundamentam-se os conceitos básicos das redes neurais, incluindo tipos de redes, funções de ativação de neurônios, além de apresentar as bases de funcionamento do algoritmo de aprendizagem da rede: o *backpropagation*.

Numa última etapa, no capítulo cinco, mostra-se a metodologia utilizada neste trabalho e os resultados obtidos.

2 O SISTEMA IPTV

Este capítulo partirá de uma visão macroscópica da arquitetura e funcionamento de um sistema IPTV evidenciando alguns aspectos, como protocolos transmissão utilizados, o *codec* MPEG4 utilizado na compressão do conteúdo e contará também com uma apresentação dos elementos que compõem o sistema (serviços oferecidos).

Será destacado aqui o ponto na rede em que poderá ser aplicada a melhoria proposta.

2.1 ARQUITETURA DE REFERÊNCIA IPTV

Existe uma grande demanda por uma padronização do sistema IPTV e de que forma este serviço deve ser oferecido aos usuários. Porém, prevalecem argumentos que tal serviço deve ser oferecido numa plataforma baseada em NGN (New Generation Network). Tal esforço tem sido realizado principalmente pela ITU-T, FG IPTV, ATIS-IIF e ATIS-TMOC, ETSI TISPAN, DVB Consortium, Packet Cable e OMA BCAST, cada um trazendo características próprias em suas propostas [5].

Existem características em comum básicas entre os modelos genéricos NGN e IPTV que viabilizam a integração do último ao primeiro, mas não são só questões técnicas que viabilizam este aproveitamento.

Em [6] são destacadas três opções de arquitetura de rede: utilizar a rede já existente e consolidada e trabalhar em termos de desempenho de rede (QoS) para que o tráfego IPTV atenda às expectativas do cliente; construir uma rede totalmente nova e dedicada à nova aplicação; ou dedicar canais exclusivos para o tráfego IPTV na rede. É enfatizado então, mais uma vez, que do ponto de vista econômico também é vantajoso aproveitar ao máximo as facilidades que o modelo NGN genérico oferece atreladas à técnicas de transmissão *multicast* e utilizar múltiplos servidores a fim de atender a demanda VoD (*Video on Demand*).

Desta forma, apresentaremos aqui uma arquitetura mais geral de um sistema IPTV, desconsiderando particularidades das inúmeras propostas de padronização existentes, mas não desprezando o fato de que as aproximações

tecnológicas dos modelos NGN e IPTV favorecem a integração de serviços. O objetivo aqui é situar o leitor e chamar atenção para o ponto específico em que está sendo aplicada a pesquisa. A figura 1 mostra os elementos que compõem uma arquitetura de rede IPTV baseada em acesso ADSL.

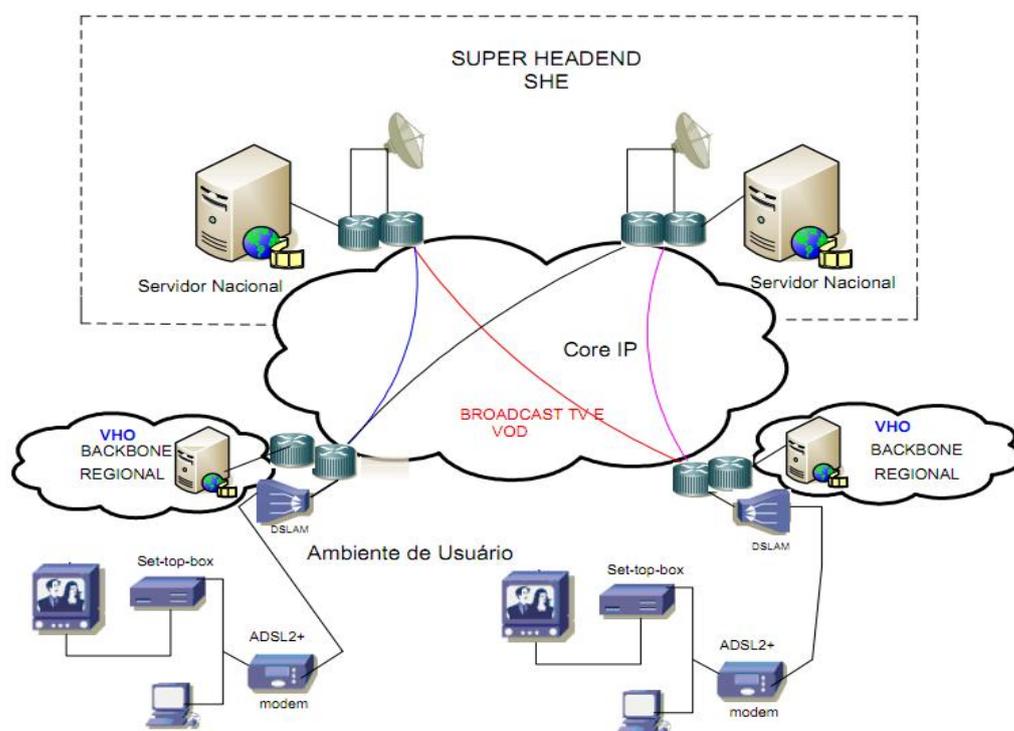


Figura 1: Arquitetura e elementos IPTV

2.2 HEADEND

O *headend* é constituído de vários elementos, que variam de rede para rede. É aqui que se diferencia a maioria das propostas pela padronização do sistema IPTV.

De maneira geral, no *headend* estão incluídos os codificadores e decodificadores, todo o sistema de controle de acessos de usuários, os servidores centrais de conteúdo, além das conexões com os provedores de conteúdo, multiplexadores, moduladores e encoders [7]. Ou seja, toda a gerência do tráfego é realizada numa estrutura encontrada no SHE.

A figura 1 ilustra uma arquitetura descentralizada de *headend*, onde são distribuídos VHOs (*Video Hub Offices*) com a finalidade de reduzir os atrasos de

rede entre o SHE (*Super Headend*) e o usuário final [6]. Para isso, cada VHO tem a função de armazenar vídeos mais demandados e conteúdo local, como propagandas e programação regional. Normalmente são fornecidas pelo menos duas conexões com a rede IP para garantir disponibilidade de serviço em caso de falha de um dos servidores.

O vídeo é codificado e processado (MPEG4) no SHE, para depois ser entregue ao *backbone* IP, onde o sinal é encapsulado com a utilização do protocolo IP e distribuído aos usuários.

A compressão do vídeo é um dos requisitos básicos para que o sistema IPTV funcione, pois isto garante uma enorme economia de banda na transmissão.

2.2.1 COMPRESSÃO DE VÍDEO MPEG4 (MOVING PICTURE EXPERT GROUP)

Existem duas maneiras de comprimir dados: com perdas ou sem perdas. Um sistema de compressão sem perdas possibilita a reconstrução fiel do sinal original após o processo de descompressão. Porém este tipo de sistema alcança um nível modesto de compressão, então utiliza-se de métodos com perdas a fim de se atingir altos níveis de compressão, como é o caso do MPEG.

A compressão de um sinal que carrega a informação pode ser realizada eliminando redundâncias do sinal. Este é o princípio básico da compressão.

Existem dois tipos de redundância de sinal que são consideradas para compressão de vídeo: a espacial e a temporal. A primeira diz respeito à repetição de informações contidas num mesmo quadro de vídeo, enquanto a segunda indica “o que muda na imagem” de um quadro para o outro.

Assim, existem duas técnicas de codificação da imagem. A técnica Intra quadro se refere à codificação realizada dentro de um mesmo quadro, não havendo relação com os outros quadros da sequência de vídeo. Já a técnica Inter quadro é a técnica que proporciona as maiores taxas de compressão, pois é ela que explora as redundâncias temporais do vídeo. A técnica de compressão temporal do padrão MPEG é baseada na estimação de movimento.

A figura 2 mostra uma sequência de fotos tiradas de um *skatista*, onde podemos observar que o único objeto que se move é o rapaz que realiza a

manobra. Cada repetição do garoto na figura representa um quadro, sendo o primeiro codificado com a técnica Inter quadro e todos os outros subsequentes com a técnica Intra quadro.



Figura 2 – Sequência de fotos

O que se faz para comprimir vídeo é transmitir somente a diferença entre os quadros subsequentes, pois a semelhança entre o conjunto de imagens que forma o vídeo é muito grande.

A técnica temporal define três tipos de quadros: os quadros I, P e B.

Quadro I: é o quadro principal, que recebe codificação Intra quadro, ou seja, não depende dos outros quadros da sequência do vídeo.

Quadro P: é previsto a partir de um quadro I ou outros quadros P da sequência de quadros codificado. É gerado pelo quadro imediatamente anterior a ele e contém apenas informações relativas à diferença entre os quadros.

Quadro B: é bidirecional, ou seja, é gerado a partir de um quadro P ou I, sejam eles posterior ou anterior ao B. O grupo de figuras (GOP – *Group of Pictures*) é definido pela sequência de quadros I, B e P até a repetição do quadro I. Um GOP é mostrado na figura 3.

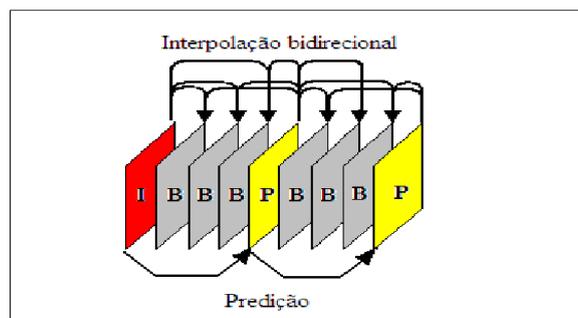


Figura 3 – Group of Pictures (MPEG4)

Tendo em vista esta correlação entre quadros, observa-se a grande importância do quadro I na transmissão de um vídeo codificado através do padrão MPEG, pois os outros dois dependem deste para existir.

O H.264/AVC (*Advanced Video Coding*) [2,3,8], mais conhecido como MPEG4, é o mais novo padrão de vídeo desenvolvido em conjunto pelos grupos internacionais de estudos VCEG da ITU-T e MPEG da ISSO/IEC, atingindo taxas de compressão até 50% maiores que o MPEG2 lançado anteriormente.

2.3 NÚCLEO DA REDE

O conteúdo de vídeo é transportado sobre uma estrutura física denominada núcleo da rede. É normalmente constituído por uma estrutura baseada em fibra óptica (que pode utilizar tecnologia de multiplexação como DWDM) e roteadores.

Aqui são contemplados protocolos de transporte, sinalização e *multicast*.

Os roteadores são limitados no sentido de capacidade de fluxo de pacotes. Quando o fluxo de pacotes se torna demasiadamente intenso, ultrapassando a capacidade nominal de roteamento (em bits/s) e de *buffer* do roteador, ocorrem descartes de pacotes.

No caso do fluxo de vídeo IPTV sendo transportado pela rede, não é aceitável que pacotes I sejam descartados neste processo, pois isto degradaria o serviço por alguns segundos no ambiente do cliente. Por isso é desejável que sejam implementados mecanismos de descarte seletivo nos roteadores, de forma a garantir parâmetros de qualidade ao usuário.

2.3.1 PROTOCOLOS DE TRANSPORTE

Os protocolos de transporte numa rede IP são basicamente o TCP e o UDP. A diferença principal entre os dois é que o UDP não garante entrega de pacotes, uma vez que não há reenvio e nem controle de recebimento de pacotes. Já o TCP trabalha com garantia de recebimento, reenvio de pacotes que por algum motivo não chegam ao seu destino e controle de fluxo de dados enviados.

Uma vez que o reenvio de pacotes não faz sentido em aplicações em tempo real, como é o caso do IPTV, neste caso é utilizado o protocolo UDP.

Protocolo UDP

Como já comentado no texto, o protocolo UDP é utilizado em comunicações em que não é garantida a uma transmissão confiável, no sentido em que não há reenvio de pacotes e confirmação do recebimento destes. Porém *streams* MPEG do sistema IPTV são transportados sobre a rede IP utilizando UDP/RTP pois este possibilita um processamento mais rápido, o que reduz o tempo de resposta. A figura 4 mostra os campos do protocolo UDP.

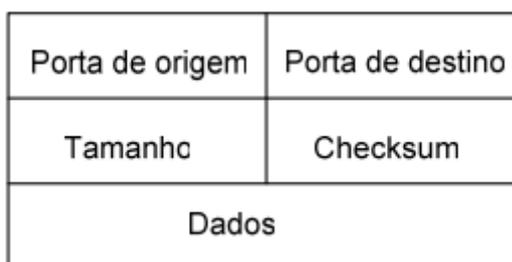


Figura 4 - Protocolo UDP

Protocolo de controle RTP

O RTP (*Real-Time Transport Protocol*) foi apresentado em 1996 através da RFC 1889 [9] com o objetivo de padronizar funcionalidades para os aplicativos de transmissão de dados em tempo real, mas sem garantia de QoS ou reserva de recursos da rede.

É um protocolo que opera sobre UDP/IP, de forma que o conteúdo (*payload*) é encapsulado em RTP, que por sua vez, é encapsulado num segmento UDP. O datagrama do protocolo RTP é mostrado abaixo.

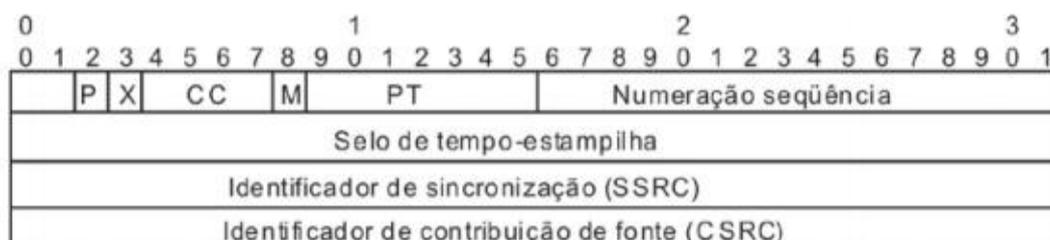


Figura 5 - Protocolo RTP

O RTP é visto pela camada de transporte UDP como parte do *payload*, como mostra o datagrama de um conteúdo encapsulado em RTP/UDP/IP.

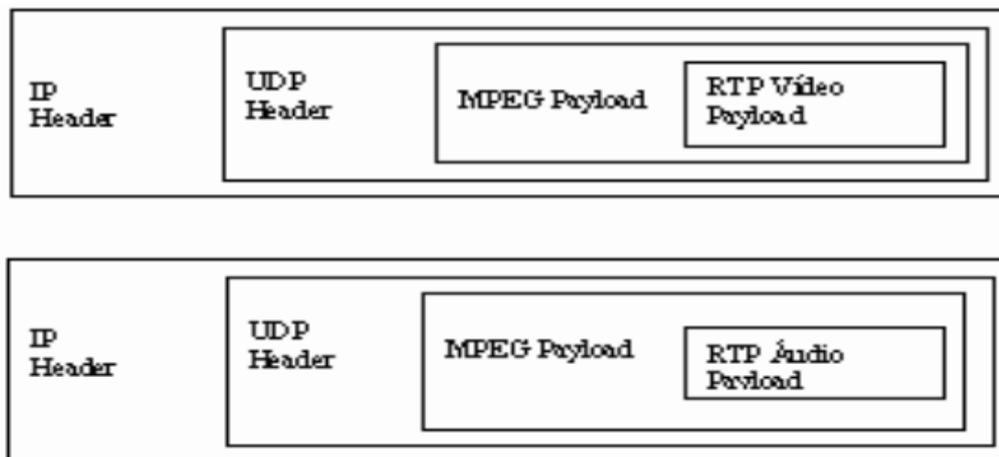


Figura 6 - Encapsulamento

3 MECANISMOS DE GERÊNCIA DE FILAS

O controle de tráfego de pacotes que entram e saem por um nó da rede é realizado pelo conjunto de sistemas de filas e mecanismos associados que permitem que o roteador tome decisões com relação ao que fazer com os pacotes que chegam a ele.

Em mecanismos chamados *Classfull* [10], fluxo de pacotes chega ao roteador e passa por uma classificação, de acordo com a sua aplicação (voz, vídeo, dados, etc.) ou outros parâmetros, dependendo do tipo. O classificador permite que o tráfego de diferentes aplicações seja reconhecido e assim processado em fluxos separados.

Porém, normalmente os mecanismos *classfull* são associados a mecanismos *classless*, como o FIFO (*First in First Out*) e RED (*Random Early Detection*) para que juntos possam compor o sistema de controle de tráfego.

3.1 FIFO

Este é o mecanismo de escalonamento de pacotes mais simples e trata de uma única fila de maneira que os pacotes não são diferenciados e vale a regra básica de que sai antes quem chega primeiro (*First in First Out*).

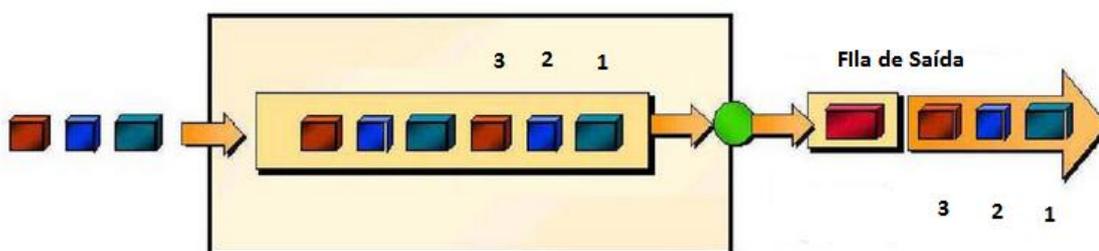


Figura 7 – FIFO

A simplicidade e eficiência estão entre as maiores vantagens deste mecanismo, sendo que é um algoritmo que não sobrecarrega os roteadores.

Quando a capacidade da fila chega a seu limite o FIFO utiliza a técnica *drop-tail*, que consiste em descartar os pacotes excedentes à capacidade da fila do roteador. O *drop-tail* pode levar a problemas no fluxo da rede porque não

permite realizar a priorização do descarte, de modo que pacotes importantes podem ser descartados .

No caso do IPTV a desvantagem é evidente, pois o tráfego é feito em rajadas e caso um pacote tipo I seja descartado neste processo, os próximos frames (B e P) não poderão ser resolvidos e a imagem final será degradada.

3.2 RED

O *Random Early Detection* apresenta-se como uma melhoria ao FIFO. Esta melhoria consiste em enviar uma resposta a fluxos que implementem controles de congestionamento antes que a fila encha. A ideia é indicar que a congestão está eminente através de uma comparação entre o máximo e o mínimo limiar e o tamanho médio das filas nos roteadores.

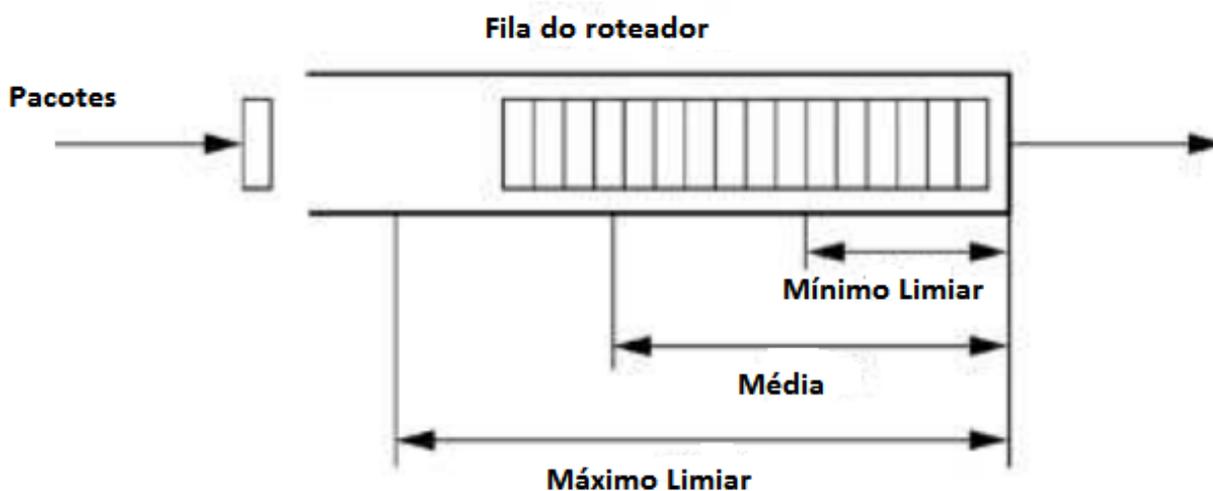


Figura 8 - RED

Quando o tamanho médio da fila é menor que o limiar mínimo, nenhum pacote é marcado para descarte. Porém quando o tamanho médio é maior que o máximo limiar, todos os pacotes subsequentes são marcados para descarte. Valores intermediários introduzem diferentes probabilidades de o pacote ser descartado [10].

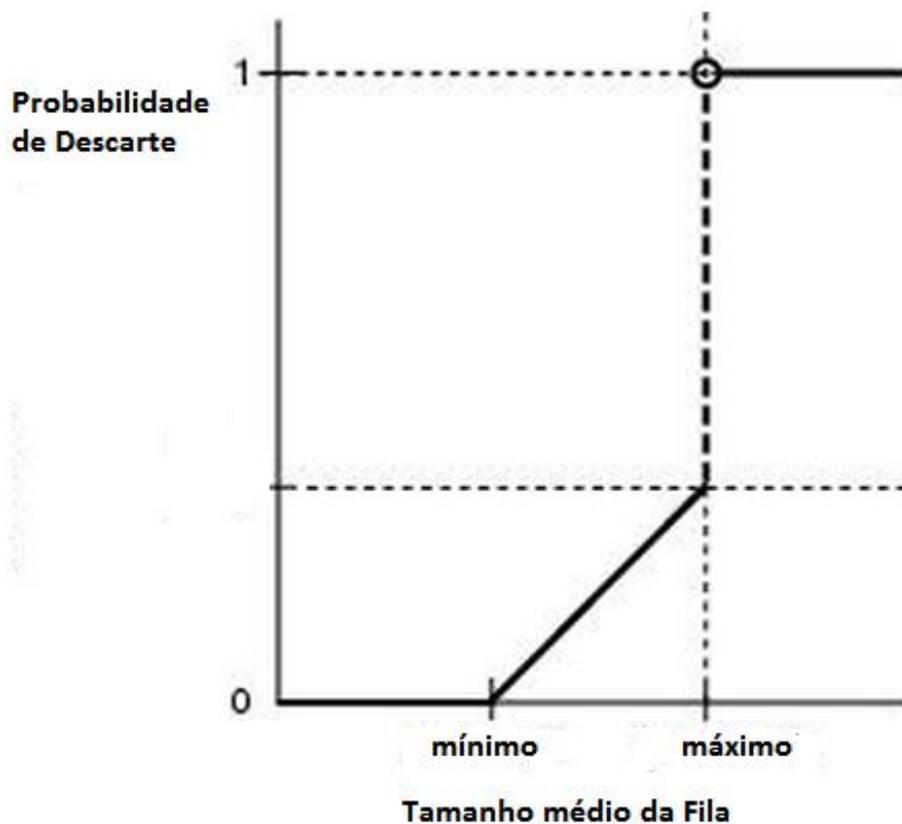


Figura 9 - Probabilidade de Descarte de Pacotes RED

Os mecanismos de escalonamento, como já comentado, são normalmente combinados para se obter o efeito desejado, como mostra figura 10. Desta forma podem ser implementados inúmeros algoritmos escalonadores, como o SFQ, WFQ, entre outros que não serão descritos neste trabalho. A figura 10 representa a combinação destes algoritmos implementados em um roteador e indica para onde o objetivo específico deste trabalho está voltado: o reconhecimento da carga transportada nos pacotes de cada uma das filas, de modo a estabelecer um mecanismo de descarte que preserve a QoS no caso de congestionamento.

4 REDES NEURAIS ARTIFICIAIS

4.1 PRINCÍPIOS DE FUNCIONAMENTO

Uma rede neural artificial é uma rede computacional que surgiu com intuito de funcionar de forma análoga ao cérebro. A ideia principal é poder reproduzir artificialmente a capacidade de aprendizagem do neurônio biológico. As máquinas não devem mais simplesmente responder de acordo com que foram previamente programadas, mas sim adaptar sua programação conforme o estímulo que recebem. As semelhanças entre o funcionamento artificial e o biológico encontram-se em dois aspectos principais, conforme é citado em “Redes Neurais”, de Simon Haykin, 2001 [11]: a experiência é fonte do conhecimento adquirido, enquanto este é armazenado nas sinapses.

A célula nervosa do cérebro responsável por captar os estímulos externos, transmiti-los e processá-los é o neurônio. Ele é constituído por um corpo celular, denominado “soma”, dendritos e axônios, conforme mostra a figura 11. O corpo celular, onde encontra-se o núcleo da célula, é responsável pelo processamento do sinal recebido pelo dendrito, que recebe os sinais de estímulo, sejam estes externos ou provenientes de outro neurônio. O axônio é responsável pela transmissão do sinal processado ao neurônio seguinte.

A sinapse é a conexão entre neurônios e implica na excitação ou inibição dos neurônios receptores.

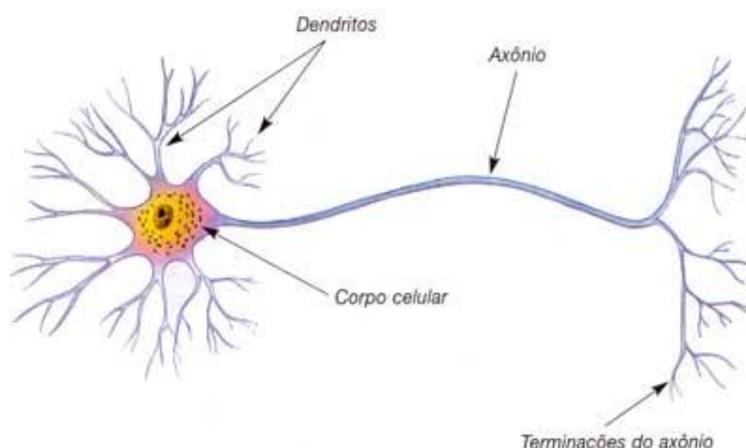


Figura 11 - Neurônio Biológico

Analogamente, em 1962, após estudos de muitos outros pesquisadores e inúmeras propostas de sistemas que funcionassem com o aprendizado baseado em reforço, Frank Rosenblatt propõe e define máquinas denominadas “*perceptrons*”. O somador e a função de ativação, mostrados na figura 12, correspondem ao *perceptron* em si, enquanto as entradas seriam os sinais de entrada e o sinal de saída, o sinal processado. O *perceptron* é então a definição de um neurônio artificial.

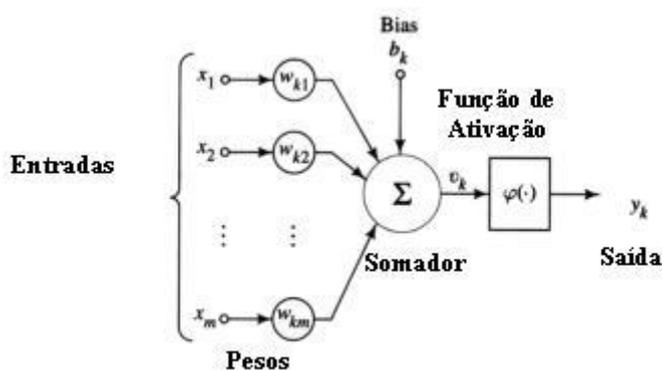


Figura 12 - Neurônio Artificial

Portanto, a saída do neurônio y_k é igual a:

$$y_k = \varphi(\sum x_m w_{km}).$$

A sinapse num neurônio artificial é caracterizada por um estímulo de entrada multiplicado pelo seu peso sináptico correspondente. O somatório destes sinais de entrada, depois de multiplicados, é aplicado a uma função de ativação que restringe a saída do sinal a valores limitados, seja de 0 a 1, ou de -1 a +1.

4.2 FUNÇÕES DE ATIVAÇÃO

A função de ativação do neurônio é o processamento do sinal que chega através das sinapses e deve simular as características não lineares do neurônio biológico. As funções mais utilizadas são linear e sigmóide, mas estas funções podem ser ajustadas conforme o pesquisador achar pertinente para seu objeto de

estudo. A figura 13 mostra algumas funções de ativação que poderiam ser aplicadas aos neurônios:

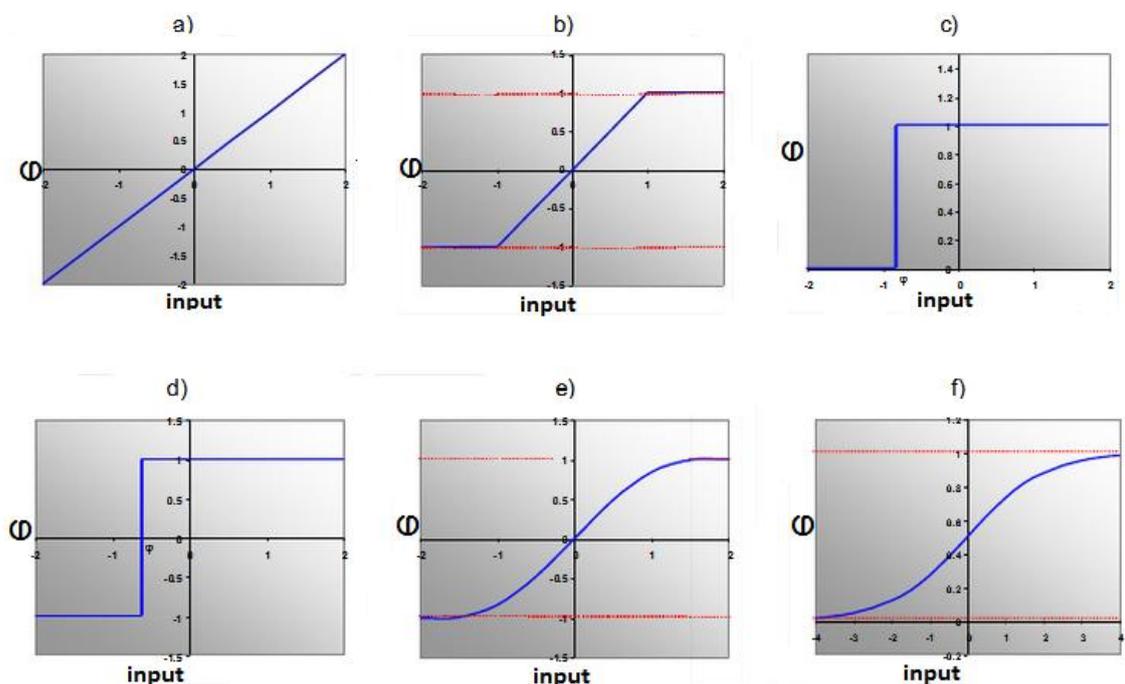


Figura 13 - Funções de ativação (a)linear; (b) linear com restrição; (c) degrau de 0 a 1; (d) degrau de -1 a +1; (e) tangente hiperbólica; (f) sigmóide

A função degrau é apropriada para sistemas binários, enquanto as funções sigmóide e tangente hiperbólica podem ser aplicadas tanto a sistemas contínuos quanto binários. A função sigmóide é a mais utilizada na literatura por apresentar comportamento similar ao neurônio biológico, além de ser contínua e apresentar derivada simples. A função sigmóide é também apropriada se utilizada nos neurônios das camadas escondidas e de contexto, pois torna o ajuste de pesos mais suave. Na saída da rede é interessante usar a função linear.

4.3 TOPOLOGIAS DE REDE

Existe uma grande variedade de redes, sendo que cada uma produz um resultado diferente. Porém não existe uma topologia soberana, que produza resultados melhores para todo e qualquer problema. Para descobrir a melhor topologia devemos testar e comparar os resultados adquiridos por cada um dos tipos de redes.

Basicamente, existem dois tipos de classificação para as redes: as alimentadas para frente (Feed-Forward) e as recorrentes, com realimentação.

As redes feed-forward se caracterizam pelo fato de o sinal se propagar somente da entrada para saída da rede, sem realimentação. As entradas de um neurônio são as saídas dos neurônios da camada anterior e não há ligação entre neurônios da mesma camada. Entre a entrada e a saída são dispostas m “camadas escondidas”, que podem conter n neurônios, conforme se desejar, porém o resultado pode variar conforme se ajustam os valores de m e n . A figura 14 exemplifica este tipo de rede, onde apresenta-se duas camadas escondidas, n elementos de entrada e p elementos de saída.

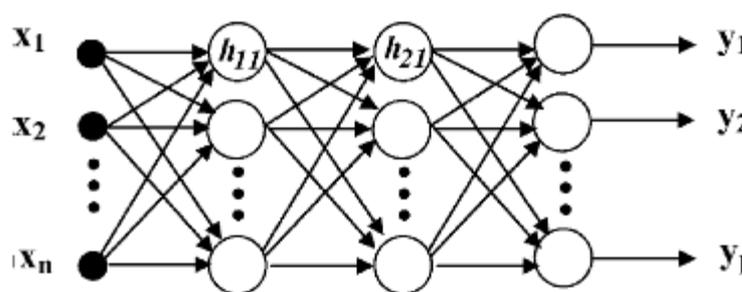


Figura 14 - Rede Feed-Forward com duas camadas escondidas

As redes recorrentes são as que apresentam realimentação, como é o caso da rede Elman, por exemplo. A rede Elman usa a realimentação da saída da camada escondida para os respectivos neurônios da “camada de contexto”, a qual realimenta todos os neurônios da camada escondida. Para tanto o número de neurônios da camada escondida deve ser igual ao da camada de contexto. A figura 15 mostra a rede Elman [12,13].

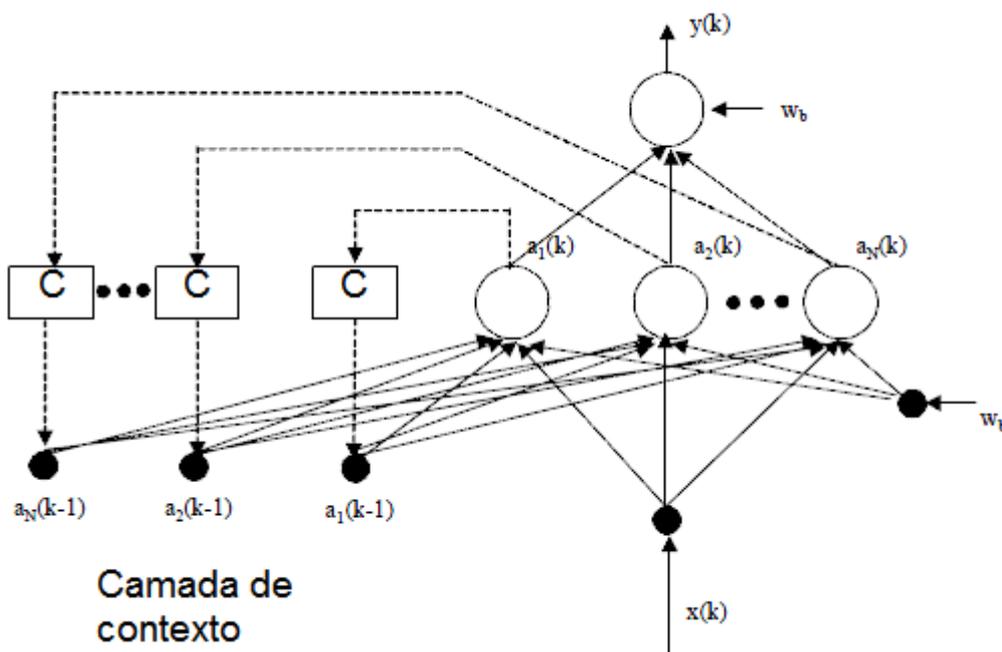


Figura 15 - Rede Elman com uma entrada e uma saída

Na figura 15, $x(k)$ representa a entrada externa; $a_i(k)$ a saída dos neurônios da camada escondida no tempo discreto k e $a_i(k-1)$, a saída dos neurônios da camada de contexto, com i variando de 1 a N . A saída da rede $y(k)$ é então representada por [12]:

$$y(k) = f(\sum w_{io} a_i(k) + w_{bo}), \text{ onde}$$

$$a_i(k) = f(\sum w_{ij} x_j(k) + \sum w_{ij} a_j(k-1) + w_{bi}).$$

W_{ij} representa os pesos sinápticos e w_b o peso *bias*, calculado através do algoritmo de aprendizagem de forma a ajustar os pesos. Desta forma, a camada de contexto funciona como uma memória para o sistema, quando a entrada da cama escondida é alimentada não só pelo sinal externo, mas também pelos valores obtidos com $x_i(k-1)$.

4.4 APRENDIZADO

O aprendizado da rede neural consiste em treiná-la para que reconheça padrões de séries temporais. Para isto são usados conjuntos de teste que

representem aquele padrão, de forma a ajustar os valores dos pesos sinápticos para valores ótimos que maximizam a assertiva da rede para outro conjunto de validação. Existem duas maneiras de aprendizado: o supervisionado e o não supervisionado.

O primeiro baseia-se num conjunto de testes, onde os valores de saída são conhecidos. Assim, a saída obtida é comparada ao valor desejado e a partir desta informação os parâmetros livres são ajustados [14]. O conjunto de testes deve ser diferente do conjunto de validação, porém deve representar o padrão da série temporal que se deseja reconhecer com a rede neural. O algoritmo utilizado neste tipo de aprendizado é o *backpropagation*. Neste trabalho, o aprendizado será do tipo supervisionado.

O segundo tipo de aprendizado, não supervisionado, é caracterizado pela ausência de um padrão de referência.

4.4.1 FUNDAMENTOS DO ALGORITMO BACKPROPAGATION

O algoritmo *backpropagation* surgiu por volta de 1974, com a publicação de Werbos [15] e a contribuição de outros pesquisadores como Rummerhart, Hinton e Williams. Este algoritmo se destacou entre outros que procuravam um valor w [$w_1, w_2, w_3 \dots w_n$] – conjunto de pesos sinápticos - que minimizasse o erro da previsão de séries, quando se utiliza redes neurais. Este era um problema fundamental para a solução dos problemas e foi o motivo de um período de estagnação nos estudos das redes neurais, até a publicação do algoritmo *backpropagation*.

O livro NSBOOK [16] nos apresenta uma dedução e demonstração deste algoritmo detalhadamente. Aqui apresentaremos os principais resultados a fim de mostrar os princípios fundamentais do algoritmo a ser utilizado no aprendizado deste estudo específico.

Tomando como exemplo um problema linear de predição dos valores de saída a partir de um conjunto de valores de entrada numa rede, num sistema de aprendizado supervisionado, temos que o erro na saída da rede é calculado comparando-se a saída que se obteve com a saída desejada, conforme abaixo:

$$J = \frac{1}{2N} \sum_{i=1}^{i=N} (y_i - w_i x_i)^2$$

onde y_i é o valor desejado da saída, w_i o peso aplicado a cada entrada x_i e N o número de padrões utilizado no aprendizado.

Deseja-se que J seja mínimo para um determinado valor de w , então o gradiente de J em relação a w deve ser nulo.

$$\mathit{grad} J = 0$$

Assim, no algoritmo *backpropagation*, busca-se o ponto da curva de erro (Figura 16) onde o erro quadrático seja nulo e para isto os valores de w_i são ajustados de forma a deslocar o ponto da curva J sempre para o lado oposto do gradiente, até atingir valor nulo.

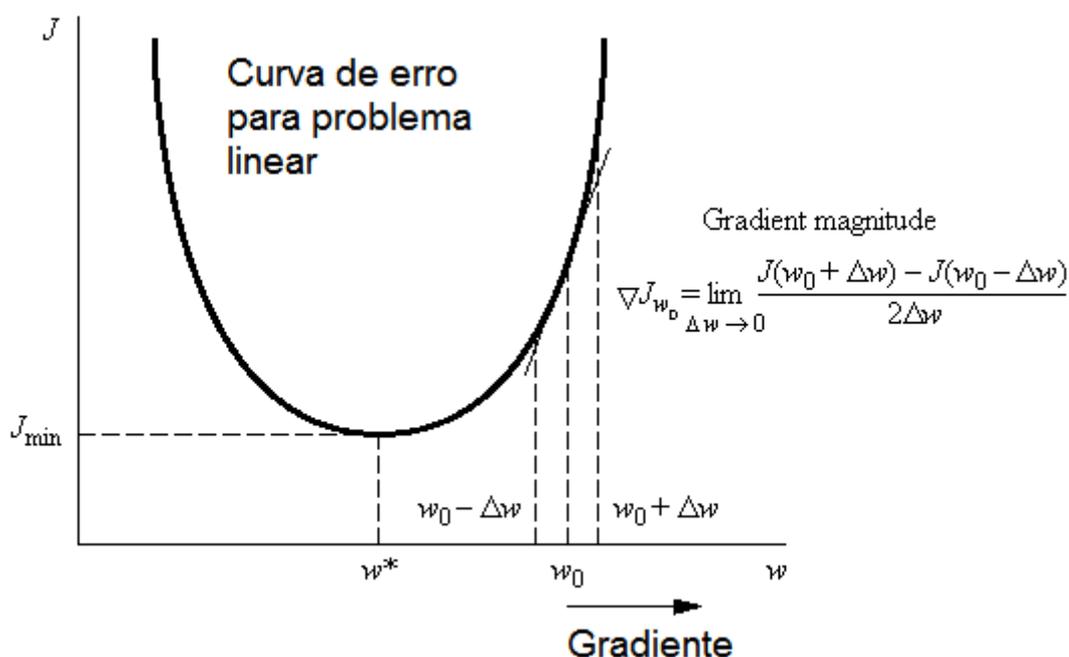


Figura 16: Curva e gradiente do erro para problemas lineares

Os valores de w para próxima iteração do aprendizado $k+1$ são então ajustados conforme equação abaixo.

$$\mathbf{w}(k + 1) = \mathbf{w}(k) - \eta \mathit{grad} J(k)$$

$$\eta \text{grad}J(k) = \Delta w$$

Quanto maior o valor do fator de aprendizado η , maior é o salto do valor de $w(k)$ para $w(k+1)$. Observando a figura 16, temos que quanto maior for η , maior será Δw de uma iteração para outra e assim o valor w^* poderá ser atingido mais rapidamente.

Em primeira vista parece vantajoso incrementar o valor do fator de aprendizagem ao máximo enquanto possível, conforme mostra a figura 17. As curvas 1, 2 e 3 representam valores diferentes para o fator de aprendizagem η , com valores crescentes de 1 a 3. Em 1 o número de iterações no aprendizado deve ser maior que em 3 para atingir o valor mínimo da função (J_{\min}).

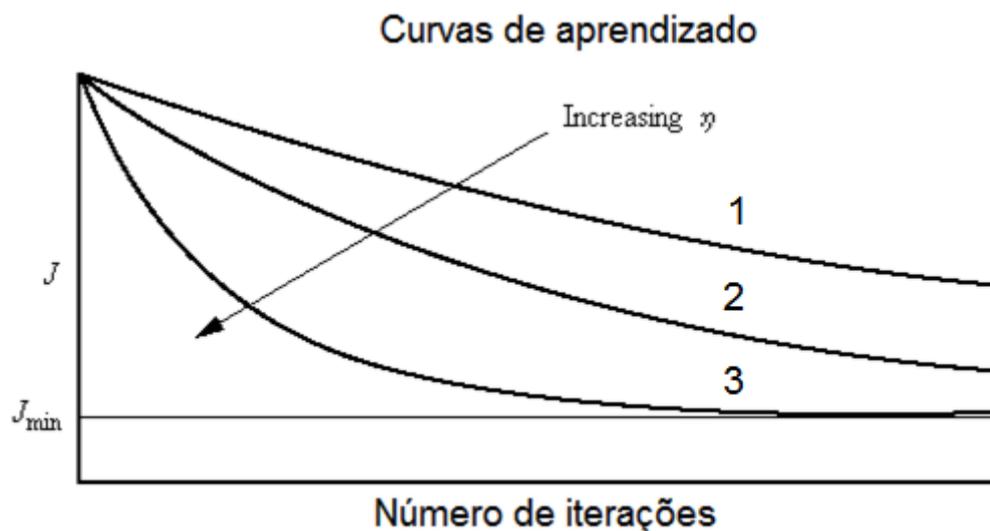


Figura 17 - Curvas de Aprendizado

Porém, deve-se tomar cuidado com estes ajustes, pois o aumento excessivo pode causar um *overflow* e o valor ótimo w^* que proporcionará um menor erro J pode nunca ser atingido.

Na primeira situação mostrada na figura 18, o valor de η proporciona um aprendizado mais lento, porém, com a possibilidade maior de atingir o mínimo da função. Na segunda situação, com valor de η maior que na primeira, o tempo de aprendizado é maior, uma vez que o salto Δw resulta em um certo “rodeio” em

torno de w^* . Na última situação, o valor de η (e conseqüentemente de Δw) é muito grande, o que resulta num número de iterações infinitas, sem atingir o valor w^* .

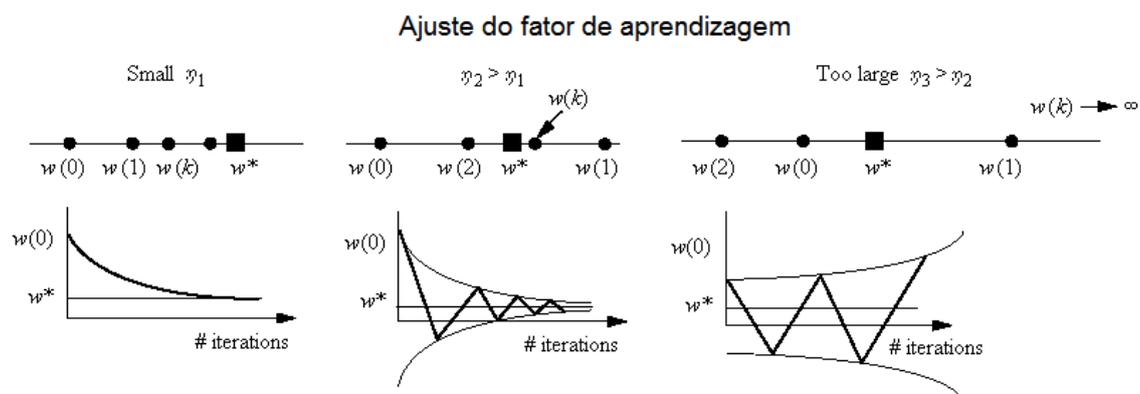


Figura 18: Ajuste do fator de aprendizagem

Deve-se levar em conta também que normalmente os problemas não são lineares, como é o caso da série temporal gerada pelo tráfego de vídeo MPEG4. Nesta situação a curva de erro não é uma parábola perfeita, mas sim uma curva com uma série de mínimos, conforme mostra a figura 19.



Figura 19 – Mínimo Local e global

Na situação de não-linearidade do problema, o algoritmo pode não encontrar o melhor valor para os pesos que satisfaçam a condição de encontrar o menor erro, pois não irá distinguir um mínimo local do global. Para contornar este problema é conveniente realizar diversos testes.

5 METODOLOGIA E RESULTADOS

Neste capítulo, numa primeira sessão será descrita a metodologia utilizada neste trabalho para na segunda sessão serem apresentados os resultados obtidos.

5.1 METODOLOGIA

Os dados para treinamento e validação das redes neurais foram extraídos de <http://www.tkn.tu-berlin.de/research/evalvid/cif.html> [17], onde contém exemplos de vídeos no formato bruto CIF (*Common Intermediate Format*) 352x288 pixels. Os exemplos utilizados para análise foram `highway_cif.yuv`; `bridge-far_cif.yuv`; `soccer_cif.yuv`; `paris_cif.yuv` e `coastguard_cif.yuv`.

A ferramenta para compressão de vídeo em formato MPEG4 foi a `ffmpeg`, disponível também em [17]. Nesta ferramenta pode-se determinar alguns parâmetros, como o tamanho do GOP e o número de quadros B entre os I, entre outros parâmetros. Para este trabalho foi selecionado GOP com tamanho 12, sendo 2 quadros B entre cada P. Assim, o GOP fica da seguinte forma: `IBBPBBPBBPBB`.

Para o treinamento da rede neural foi utilizado o método de “janelamento” [12] com 15, 25, 35, 45 e 55 pacotes na entrada das redes. Os parâmetros do pacote levados em consideração para entrada foram o tamanho e o tempo de chegada. Para cada padrão, a saída da rede deveria ser 1 se existissem pacotes pertencentes a um quadro I dentro da janela e 0 se não houvesse.

O motivo da escolha dos tamanhos das janelas é o número de pacotes por quadro. Para o formato CIF, tem-se que um quadro I possui em média 15 pacotes, enquanto os quadros B e P, 10 pacotes. Desta forma o GOP possui em média 125 pacotes e a janela com 55 entradas estaria abrangendo quase metade do GOP. Assim, quanto maior a janela em teste, maior a probabilidade de ali haver pacotes pertencentes ao quadro I.

O parâmetro de comparação foi a porcentagem de erros das topologias em teste com relação a saída da rede. Foi comparado o desempenho de uma rede

FeedForward e uma recorrente Elman, ambas utilizando o algoritmo de aprendizagem *backpropagation*, com fator de aprendizagem 0,2 e 50000 ciclos para treinamento da rede. O software utilizado foi o JavaNNS, que pode ser obtido em <http://www.ra.cs.uni-tuebingen.de/downloads/JavaNNS/>.

A figura 20 ilustra o JAVANNS com as janelas utilizadas para monitoramento de testes. Acima e à esquerda encontra-se a rede com seus neurônios e sinapses, sendo cada um passível de configuração da função de ativação. No canto direito superior, as curvas de erro de treinamento e validação, em preto e vermelho, respectivamente. As janelas inferiores correspondem ao painel de controle, onde são setados os parâmetros do algoritmo de treinamento, e ao campo de monitoramento dos *logs*.

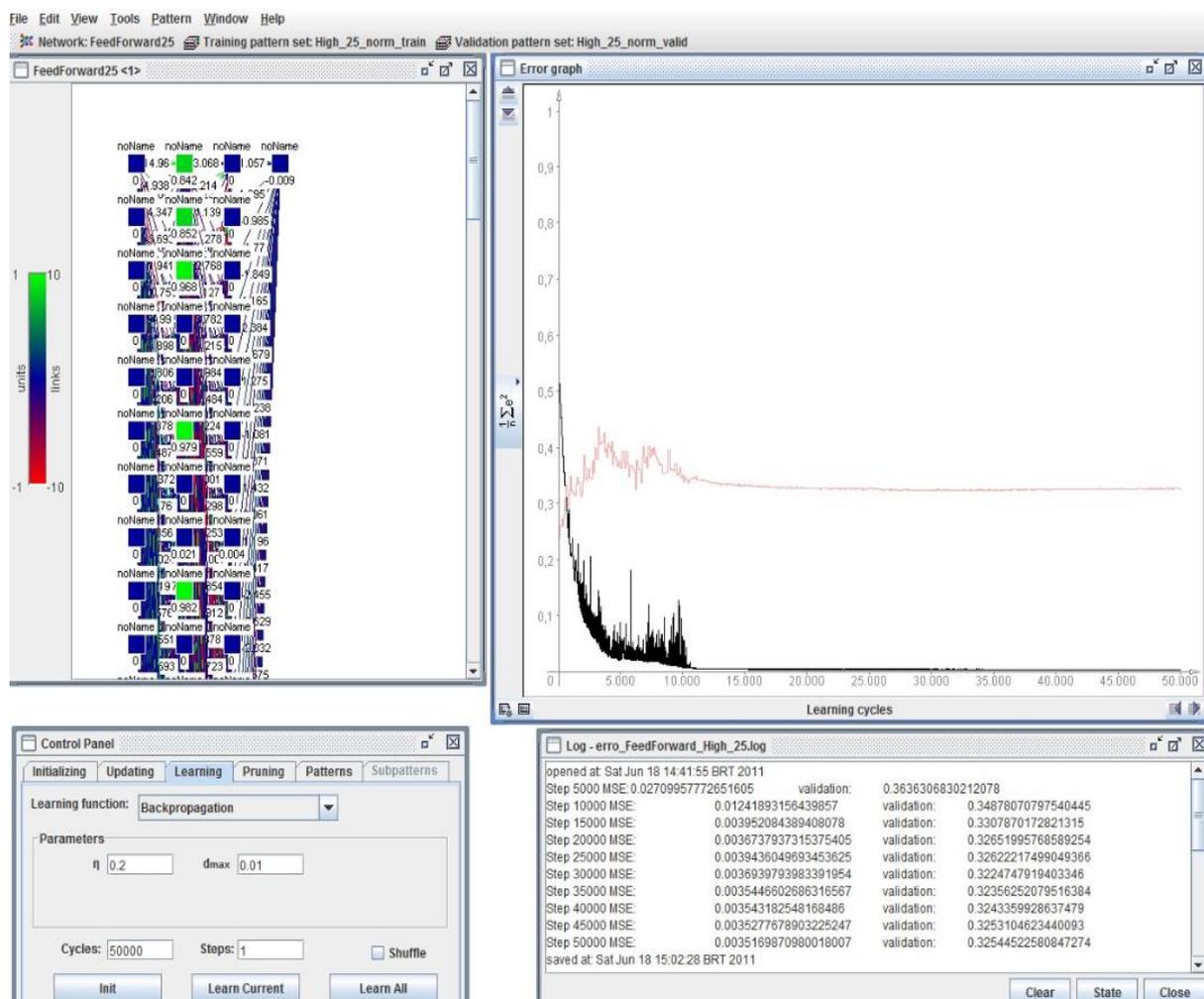


Figura 20 - JavaNNS

5.2 RESULTADOS

As tabelas 1 a 5 apresentam os erros obtidos na saída das redes para padrões com 15, 25, 35, 45 e 55 pacotes na entrada (janelas), respectivamente. Foram separadas as estatísticas de assertiva de padrões que contém pacotes pertencentes ao quadro I dos que não contém. Todos os valores das tabelas são em relação ao número total de padrões testados, para cada vídeo.

Janela 15	Elman		FeedForward	
	Presença de quadro I não reconhecida	Ausência de quadro I não reconhecida	Presença de quadro I não reconhecida	Ausência de quadro I não reconhecida
Bridge	10,50%	2,30%	38,30%	10,30%
Coast	34,80%	5,10%	78,30%	33,10%
Highway	7,70%	1,20%	50,00%	16,00%
Paris	2,10%	0,40%	4,80%	6,60%
Soccer	10,00%	0,90%	50,00%	16,20%

Tabela 1

Janela 25	Elman		FeedForward	
	Presença de quadro I não reconhecida	Ausência de quadro I não reconhecida	Presença de quadro I não reconhecida	Ausência de quadro I não reconhecida
Bridge	2,40%	0,30%	1,00%	0,60%
Coast	15,20%	0,00%	9,10%	9,00%
Highway	3,90%	0,00%	1,10%	0,90%
Paris	0,70%	0,00%	0,00%	0,00%
Soccer	9,40%	0,00%	0,00%	0,00%

Tabela 2

Janela 35	Elman		FeedForward	
	Presença de quadro I não reconhecida	Ausência de quadro I não reconhecida	Presença de quadro I não reconhecida	Ausência de quadro I não reconhecida
Bridge	1,10%	0,50%	0,00%	0,00%
Coast	14,30%	3,40%	0,00%	0,00%
Highway	1,90%	0,00%	0,00%	0,00%
Paris	0,00%	0,00%	0,00%	0,00%
Soccer	3,80%	0,00%	0,00%	0,00%

Tabela 3

Janela 45	Elman		FeedForward	
	Presença de quadro I não reconhecida	Ausência de quadro I não reconhecida	Presença de quadro I não reconhecida	Ausência de quadro I não reconhecida
Bridge	0,60%	0,00%	0,00%	0,00%
Coast	6,90%	2,60%	0,00%	0,00%
Highway	0,70%	0,00%	0,00%	0,00%
Paris	0,00%	2,90%	0,00%	0,00%
Soccer	0,00%	0,00%	0,00%	0,00%

Tabela 4

Janela 55	Elman		FeedForward	
	Presença de quadro I não reconhecida	Ausência de quadro I não reconhecida	Presença de quadro I não reconhecida	Ausência de quadro I não reconhecida
Bridge	0,00%	0,00%	0,00%	1,10%
Coast	7,10%	0,00%	3,60%	7,10%
Highway	1,40%	2,00%	0,00%	0,00%
Paris	0,00%	0,00%	0,00%	0,00%
Soccer	0,00%	0,00%	0,00%	0,00%

Tabela 5

Observa-se que para padrões com 25 ou mais pacotes na entrada, a rede Feed-Forward mostra um desempenho semelhante ou até melhor no aprendizado, em comparação à rede Elman. Porém as diferenças não são grandes. Entretanto, para janela com 15 pacotes na entrada, a rede Elman apresenta uma resposta muito melhor que a rede Feed-Forward.

O vídeo *Coastguard* dificulta o aprendizado das redes, o que pode ser percebido nas tabelas 1 a 5, onde ambas as redes apresentam valores maiores de erro para ele. O motivo disto é a movimentação de imagens contida no vídeo, que aumenta a não-linearidade dos dados.

6 CONCLUSÕES

Para todos os vídeos em teste, independente do tamanho da janela de pacotes utilizada, ambas as redes aprenderam o padrão, cada uma apresentando uma taxa de erro com relação ao número de acertos por padrão (janela). Quanto menor este erro significa que a rede em teste aprendeu melhor o padrão.

Apesar dos índices de erro muito baixos para padrões com 25 ou mais pacotes na entrada (tabelas 2 a 5), deve-se levar em consideração a notoriedade da melhor performance da rede Elman em relação a rede Feed-Forward no reconhecimento da presença de pacotes l em janela reduzida, com 15 pacotes na entrada (tabela 1). A explicação para isto está no fato de ser mais interessante reconhecer a presença de tais pacotes de forma mais precisa, sendo que o ideal seria identificar pacote a pacote. Além disto, os erros obtidos nas redes Elman e Feed-Forward para janelas maiores não apresentam diferenças grandes, o que dificulta uma comparação entre o desempenho das duas.

Portanto, analisando os resultados obtidos, recomenda-se utilizar a rede Elman para janelas menores, onde o desempenho entre as redes torna-se evidente.

Para trabalhos futuros, pode-se gerar o algoritmo a partir da rede treinada, testar desempenho com tráfego de vídeo MPEG4 semelhante e testar o desempenho e qualidade do vídeo recebido, além de ser possível realizar testes com vídeos em resoluções mais altas.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] MARPE, Detlev et al. **The H.264/MPEG4 Advanced Video Coding Standard and its Applications**. IEEE Communications Magazine, p. 134-143, ago. 2006.
- [2] RICHARDSON, Iain E. G. **H.264 and MPEG-4 Compression**. Video Coding for Next-generation Multimedia. Disponível em: <<http://mirror05.x264.nl/Dark/videobook.pdf>>. Acesso em: 14 fev. 2011.
- [3] ALBUQUERQUE, Marcio Portes de et al. **O padrão MPEG para compressão de Vídeo**. Disponível em: <<http://biblioteca.cat.cbpf.br/pub/apub/nt/2000/nt00200.pdf>>. Acesso em: 02 maio 2011.
- [4] TELECO, Inteligência em Telecomunicações. Disponível em: <<http://www.teleco.com.br/IPTV%20mundo.asp>>. Acesso em: 11 maio 2011.
- [5] VOLK, Mojca et al. **Quality-Assured Provisioning of IPTV Services within the NGN Environment**. IPTV Systems, Standards and Architectures: Part II. IEEE Communications Magazine, p. 118-126, maio 2008.
- [6] CHA, Meeyoung et al. **Case Study: Resilient Backbone Design for IPTV Services**, IPTV Workshop, International World Wide Conference, maio 2006, Edinburgh, Scotland, United Kingdom.
- [7] DUQUE, Luciano Henrique. **Avaliação da Qualidade de Vídeo em Redes IPTV com Acesso Baseado em ADSL**. Dissertação Mestrado (Engenharia Elétrica) – Universidade de Brasília, Brasília, 2008.
- [8] TORRES, Daniel Pioli. **Impacto do Gerenciamento Ativo de Filas em Redes IP Transportando Fluxos de Vídeo H.264/MPEG-4 AVC**. Dissertação Mestrado (Engenharia Elétrica) – Universidade Federal do Paraná, Curitiba, 2009.
- [9] SCHULZRINNE, H. et al. **RTP: A Transport Protocol for Real-Time Applications**. Network Working Group. Jan. 1996.
- [10] LIMA, Luísa; VILELA, João. **Relatório - Mecanismos de escalonamento de pacotes**. Jun. 2004. Disponível em: <<http://www.dcc.fc.up.pt/~pbrandao/aulas/0405/TAR/relatorios/0304-t2-scheduling.pdf>>. Acesso em: 02 abr. 2011.
- [11] HAYKIN, Simon. **Redes Neurais. Princípios e prática**. Porto Alegre, RS: Bookman, 2001.

- [12] CARVALHO, Janaina Veiga; ROMANEL, Celso. **Redes neurais temporais aplicadas ao monitoramento de barragens**. RESI- Revista eletrônica de Sistemas de Informação, ed. 10, n. 1, 2007.
- [13] ABDENNOUR, Adel. **Evaluation of Neural Network Architectures for MPEG-4 Video Traffic Prediction**. IEEE Transactions On Broadcasting, v. 52, n. 2, p. 184-192, jun. 2006.
- [14] Redes Neurais na Previsão de séries temporais. Disponível em: <<http://inf.unisul.br/~ines/workcomp/cd/pdfs/2878.pdf>>. Acesso em: 04 abr. 2011.
- [15] ABELÉM, Antonio Jorge Gomes. **Redes Neurais Artificiais na Previsão de Séries Temporais**. Dissertação Mestrado (Engenharia Elétrica) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 1994.
- [16] Jose C. Principe, Neil R. Euliano, W. Curt Lefebvre, Neural and Adaptive Systems: Fundamentals Through Simulations, John Wiley and Sons, 2000
- [17] KLAUE, Jirka. **EvalVid - A Video Quality Evaluation Tool-set**. Disponível em: <<http://www.tkn.tu-berlin.de/research/evalvid/>>. Acesso em: 22 mar. 2011.
- [18] BULLINARIA, John. **Quick Guide to javaNNS**. Disponível em: <<http://www.cs.bham.ac.uk/~jxb/NN/javaNNS/javaNNSguide.html>>. Acesso em: 03 fev. 2011.
- [19] PIRES, Daniel Fernandes Theophilo de Almeida. **Proposta de um Sistema de Distribuição de Licenças de Uso para IPTV**. Dissertação Mestrado (Engenharia) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2007.