

Avaliação de riscos em projetos de software a partir do uso de técnicas de inteligência computacional

Trabalho de Conclusão de Curso Engenharia da Computação

Raphael José D'Castro

Orientador: Profa. Cristine Gusmão

Raphael José D'Castro

Avaliação de riscos em projetos de software a partir do uso de técnicas de inteligência computacional

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia da Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Recife, junho de 2009.

Aos que buscam um mundo melhor.

Agradecimentos

A realização deste trabalho só foi possível em todos os seus aspectos graças as mais variadas contribuições. Porém é necessário destacar a dedicação da minha orientadora, a Profa. Cristine Gusmão, que além de oferecer todo o apoio necessário foi a maior entusiasta deste estudo. Aos Colegas Kenelly, Mateus, Wilmar e outros do PROMISE que me acolheram e ajudaram incondicionalmente. Ao Prof. Mêuser sempre atencioso e pronto para contribuir. Também não poderia deixar de agradecer a Marina, a minha família e meus amigos que são fundamentais em tudo que faço.

Resumo

O relatório *Chaos Report* apresentado no ano de 2009 mostra que apenas 32% dos projetos de software são finalizados com êxito. Estes dados mostram que décadas após a crise do software muito ainda deve ser feito. Neste período a sistematização e aplicação de boas práticas e metodologias vêm mudando a realidade, porém, este avanço tem sido lento devido as dificuldade das pequenas organizações em implantar processos. O gerenciamento de risco é um caso emblemático. Existe um consenso da importância desta disciplina, porém existem poucas ferramentas e técnicas disponíveis para oferecer suporte aos gestores. Além da escassez, as ferramentas disponíveis são oferecidas a valores proibitivos para a realidade da maior parte das organizações. A transformação da situação destas organizações passa por uma mudança interna. Atualmente o registro dos dados dos projetos não é priorizado e, desta forma, a aprendizagem com a própria experiência fica comprometida. Com base nestes dados acumulados este trabalho estuda a relação entre as características dos projetos e os resultados destes com o intuito de efetuar previsões sobre os projetos futuros. As previsões são efetuadas a partir do treinamento de uma Rede Neural Artificial *Multilayer Perceptron* para cada pilar do gerenciamento de risco. Os resultados obtidos na previsão de finalização do projeto, sob a visão de prazo e custo foram 62,5% de acerto. Para escopo a previsão atingiu resultado de 75%. Estes resultados indicam a viabilidade da utilização de Redes Neurais Artificiais para construção de ferramentas de suporte aos gestores.

Abstract

According to the Chaos Report, presented in 2009, only 32% of software projects are successfully concluded. It shows that decades after the software crisis a lot must be done. In this period the systematization and application of good practices and methodologies have been changing the reality, however the progress is coming slowly due to the difficulties of small organizations to introduce process. Risk management is emblematic. Despite the importance of this discipline there are few tools and techniques available to give support to the managers. Beyond the scarcity, the available tools are too expensive for the most part of the organizations. The situation of these organizations is changing based on an internal transformation. Nowadays, as the register of the data of the projects is not prioritized, the own experience learning is compromised. Based on this accumulated data, this study verifies the relation between the project characteristics and its results with the objective of making predictions about future projects. The prediction is made through the training of an Artificial Neural Network Multilayer Perceptron for each part of risk management. The gotten results showed 62,5% of rightness. For the target, the prediction result was 75%. The results make possible the utilization of Artificial Neural Networks to build support tools for managers.

Sumário

RESUMO.....	I
ABSTRACT.....	VI
SUMÁRIO.....	VII
ÍNDICE DE FIGURAS.....	IX
ÍNDICE DE TABELAS.....	X
TABELA DE SÍMBOLOS E SIGLAS.....	XI
CAPÍTULO 1 INTRODUÇÃO.....	12
1.1 MOTIVAÇÃO.....	14
1.2 OBJETIVOS.....	14
1.3 METODOLOGIA.....	15
1.4 ESTRUTURA DO DOCUMENTO.....	16
CAPÍTULO 2 GERENCIAMENTO DE RISCOS EM PROJETO DE DESENVOLVIMENTO DE SOFTWARE.....	18
2.1 GERENCIAMENTO DE PROJETOS.....	18
2.1.1 <i>Ciclo de vida de projeto</i>	19
2.1.2 <i>Guia PMBOK</i>	19
2.1.3 <i>CMMI (SEI)</i>	20
2.1.4 <i>Normas ISO 9000-3, ISO 12207 e ISO 15504</i>	22
2.2 GERENCIAMENTO DE RISCOS.....	22
2.2.1 <i>Visão do PMBOK</i>	25
2.3 TÉCNICAS DE AVALIAÇÃO DE RISCOS.....	25
2.3.1 <i>Métodos de Identificação de Riscos</i>	26
2.3.2 <i>Análise Qualitativa de Riscos</i>	28
2.3.3 <i>Técnicas Quantitativas de Análise de Riscos</i>	28
2.4 RESUMO DO CAPÍTULO.....	29
CAPÍTULO 3 COMPUTAÇÃO INTELIGENTE.....	31
3.1 TÉCNICAS DE COMPUTAÇÃO INTELIGENTE.....	31
3.1.1 <i>Árvores de Decisão</i>	32
3.1.2 <i>Raciocínio Baseado em Casos (RBC)</i>	33
3.1.3 <i>Lógica Fuzzy (Difusa)</i>	34
3.1.4 <i>Redes Bayesianas</i>	35

3.2	REDES NEURAIAS ARTIFICIAIS	37
3.1.5	<i>MLP – Multilayer Perceptron</i>	39
3.3	RESUMO DO CAPÍTULO	41
CAPÍTULO 4 AVALIAÇÃO DE RISCOS COM BASE EM MLP		42
4.1	AVALIAÇÃO BASEADA EM DADOS PASSADOS	42
4.1.1	<i>Caracterização dos projetos (Entrada da Rede)</i>	43
4.1.2	<i>Resultados dos Projetos (Saída da Rede)</i>	44
4.1.3	<i>Coleta de Dados</i>	45
4.2	PROTOTIPAÇÃO DA MLP	46
4.2.1	<i>Tratamento e pré-processamento dos Dados</i>	47
4.2.2	<i>Topologia e parâmetro de aprendizagem da MLP</i>	48
4.2.3	<i>Resultados da MLP</i>	48
4.3	AVALIAÇÃO DOS RESULTADOS MLP	50
4.3.1	<i>Fatores positivos e negativos das técnicas</i>	51
4.4	RESUMO DO CAPÍTULO	53
CAPÍTULO 5 CONCLUSÃO E TRABALHOS FUTUROS		54
5.1	CONTRIBUIÇÃO	54
5.2	TRABALHOS RELACIONADOS	55
5.3	DIFICULDADES ENCONTRADAS	55
5.4	TRABALHOS FUTUROS	56
BIBLIOGRAFIA		57
APÊNDICE A QUESTIONÁRIO DE COLETA DE DADOS		60
APÊNDICE B E-MAIL DE DIVULGAÇÃO		61

Índice de Figuras

FIGURA 2.1. GRÁFICO DA INTENSIDADE DO PROCESSO VERSUS TEMPO PROPOSTO NO PMBOK [17].....	20
FIGURA 2.2. GRÁFICO COM RESULTADO CHAOS REPORT [22]	23
FIGURA 2.3. GRÁFICO COM UTILIZAÇÃO DAS TÉCNICAS DE IDENTIFICAÇÃO [9].....	26
FIGURA 3.1. ILUSTRAÇÃO DE UMA ÁRVORE DE DECISÃO HIPOTÉTICA [16].....	32
FIGURA 3.2. LÓGICA BOOLEANA X LÓGICA FUZZY	34
FIGURA 3.3. TOPOLOGIA COM A PROBABILIDADE DO ALARME SOA DA REDE BAYESIANA [8].....	36
FIGURA 3.4. MODELO SIMPLIFICADO DE NEURÔNIO BIOLÓGICO E PROPOSTO POR McCULLOCH E PITTS.	37
FIGURA 3.5. RNA MULTILAYER PERCEPTRON COM UMA CAMADA ESCONDIDA	40
FIGURA 4.1. PILARES DO GERENCIAMENTO DE RISCO	45
FIGURA 4.2. DISTRIBUIÇÃO DOS DADOS DE ENTRADA COLETADOS	47
FIGURA A.1. PARTE SUPERIOR DO QUESTIONÁRIO DE COLETA DE DADOS	60
FIGURA A.2. PARTE INFERIOR DO QUESTIONÁRIO DE COLETA DE DADOS	60

Índice de Tabelas

TABELA 2.1. ÁREAS DE PROCESSOS NO CMMI	21
TABELA 2.2. DESCRIÇÃO DAS TÉCNICAS DE IDENTIFICAÇÃO DE RISCOS.....	27
TABELA 2.3. DESCRIÇÃO DAS TÉCNICAS DE ANÁLISE QUALITATIVA DE RISCOS	29
TABELA 3.1. PROBABILIDADE DE O ALARME TOCAR [8]	36
TABELA 4.1. QUANTIDADE DE NEURÔNIO POR CAMADA EM CADA MLP.....	48
TABELA 4.2. APRESENTAÇÃO DOS RESULTADOS DA MLP PARA FINALIZAÇÃO.....	49
TABELA 4.3. APRESENTAÇÃO DOS RESULTADOS DA MLP PARA PRAZO	49
TABELA 4.4. APRESENTAÇÃO DOS RESULTADOS DA MLP PARA CUSTO.....	50
TABELA 4.5. APRESENTAÇÃO DOS RESULTADOS DA MLP PARA ESCOPO	50

Tabela de Símbolos e Siglas

MLP – Multilayer Perceptron

RNA – Rede Neural Artificial

RBC – Raciocínio Baseado em Casos

PMBOK – Project Management Body of Knowledge

PMI – Project Management Institute

CMM – Capability Maturity Model

SEI – Software Engineering Institute

SW-CMM – Capability Maturity Model for Software

CMMI – Capability Maturity Model Integration

SPICE – Software Process Improvement and Capability Determination

IA – Inteligência Artificial

MAPS – Modelo de Adaptação de Processo de Software

ERP – Enterprise Resource Planning (Sistema Integrado de Gestão Empresarial)

J2EE – Java 2 Enterprise Edition (Java Edição Empresarial)

TI – Tecnologia da Informação

RBF – Radial Basis Function

Capítulo 1

Introdução

Mesmo com a evolução dos computadores, das técnicas e ferramentas nos últimos anos, a produção de software confiável, correto e entregue dentro dos prazos e custos estipulados ainda é muito difícil.

Os relatórios apresentados pelo *Standish Group* (Chaos Report) nos últimos 15 anos trazem resultados nada animadores. No estudo com mais de 175.000 projetos o primeiro resultado foi publicado em 1994 [22]. Neste ano apenas 16% dos projetos estudados foram finalizados com sucesso. Quinze anos após, em 2009 o relatório apontou uma melhoria mais o resultado ainda é muito inexpressivo. Apenas 32% dos projetos foram finalizados com sucesso [22].

Mesmo os projetos cuja entrega é feita respeitando os limites de prazo e custo, possuem qualidade suspeita, uma vez que provavelmente foram feitos com muita pressão sobre os desenvolvedores, o que pode quadruplicar o número de erros de software, segundo a mesma pesquisa.

Processos orientados a documentação para o desenvolvimento de softwares, são de certo modo um limitador aos desenvolvedores. Além disso, muitas organizações não possuem recursos ou inclinação para processos de produção de software que exigem planejamento rigoroso de todos os requisitos do sistema através de documentação extensa. Por esta razão, muitas organizações, particularmente as pequenas, acabam por não usar nenhum processo, o que pode levar a efeitos desastrosos em termos de qualidade de software.

A Gerência de Riscos é uma das disciplinas mais recentes da Gerência de Projetos e sua evolução, na área de Engenharia de Software, está associada com o tratamento dos riscos nos ambientes de desenvolvimento de software. Muitos modelos e abordagens foram apresentados e utilizados nos últimos 20 anos. Contudo, uma das grandes fraquezas dessas abordagens, até o momento, foi ter negligenciado os riscos que podem surgir do relacionamento entre projetos [9]. A

utilização dos conceitos do gerenciamento de riscos em ambientes de desenvolvimento de múltiplos projetos de software reflete uma necessidade de ter os recursos disponibilizados utilizados da melhor forma.

Neste cenário as pequenas organizações são mais fragilizadas por dispor de poucos recursos. Com isto muitos dos processos, boas práticas e ferramentas desenvolvidas terminam por não ser incorporados à realidade destas organizações. Isto decorre do fato de trazer uma carga maior de trabalho em curto prazo e os benefícios aparecem a médio e longo prazo. Desta forma um requisito importante para as técnicas e ferramentas poderem ser introduzidas nos pequenos projetos é a simplicidade.

Atualmente as principais técnicas de identificação e análise de riscos utilizam-se de dados de projetos passados. Isto evidencia a importância na manutenção de um registro adequado do histórico dos projetos.

O objetivo principal das técnicas de avaliação de riscos consiste em organizar os dados acumulados com o intuito de possibilitar aprendizagem para incorporação do conhecimento em novos projetos. Muitas das técnicas existentes têm um alto grau de subjetividade envolvida uma vez que depende da interpretação dos profissionais envolvidos no processo de avaliação.

Observando todos estes aspectos que envolvem a avaliação de riscos, este trabalho vem buscar a aplicação de uma técnica que possibilite a superação das limitações existentes e de forma simples e objetiva proporcionem conhecimento para as organizações.

A computação inteligente se apresentou como uma boa alternativa para subsidiar a aprendizagem com os registros dos projetos passados. Desta forma este trabalho realizou um estudo das técnicas de Computação Inteligente com potencial para atender esta necessidade. Com base neste estudo a Rede Neural Artificial do tipo Multilayer Perceptron foi técnica escolhida para ter sua viabilidade verificada no domínio do gerenciamento de riscos.

1.1 Motivação

O cenário preocupante apresentado pelo *Chaos Report* não será mudado significativamente caso os gerentes de projetos não estejam municiados para atacar os principais problemas encontrados no desenvolvimento dos projetos de software. Atualmente as poucas ferramentas e técnicas disponíveis no mercado não conseguiram entrar no cotidiano da grande maioria das organizações.

No intuito de reverter este quadro o grupo de pesquisa PROMISE (<http://pma.dsc.upe.br/>) vem pesquisando o desenvolvimento de uma ferramenta aberta que apóie os gerentes de projetos. O *OpenmPRIME* é um projeto que se propõe a disponibilizar uma ferramenta abrangente, eficiente e de baixo custo para o gerenciamento dos riscos.

Para este projeto atingir seus objetivos é de fundamental importância a utilização de uma técnica de avaliação de riscos que favoreça a predição dos mesmos, apontando aos gerentes os riscos em potencial a partir de experiências passadas. Assim o estudo da aplicação de uma rede *Multilayer Perceptron* – MLP para este fim se mostra bastante oportuno.

1.2 Objetivos

O gerenciamento de riscos em projetos de software tem como principal pilar de sustentação a experiência acumulada por especialistas. Estes profissionais ao longo dos anos de trabalho na área acumularam conhecimento suficiente para identificar as condições que produzem os riscos.

Com base nestas idéias este trabalho busca verificar a viabilidade de extrair estas informações a partir da busca por padrões de comportamento nos projetos. Para isto uma Rede Neural Artificial será treinada e testada visando avaliar a eficiência da Rede em realizar predições.

Além do propósito geral apresentado este trabalho tem os seguintes objetivos específicos:

- a) Coletar dados significativos sobre projetos passados com o intuito de subsidiar o treinamento de uma Rede Neural Artificial;
- b) Propor uma topologia de Rede Neural Artificial que apresente resultados satisfatórios visando o objetivo geral deste trabalho;
- c) Realizar o treinamento supervisionado da Rede Neural Artificial proposta;
- d) Verificar a eficácia da Rede Neural Artificial na predição de resultados;
- e) Comparar os resultados encontrados com a utilização de outras técnicas de computação inteligente para avaliação de riscos em projeto de software.
- f) Subsidiar o OpenmPRIME com uma técnica para avaliação de riscos.

1.3 Metodologia

Na busca dos objetivos levantados este trabalho foi organizado em quatro etapas: (i) Estudo de avaliação de riscos, (ii) Estudo de técnicas de computação inteligente na avaliação de riscos, (iii) Estudo e aplicação de redes MLP na avaliação de riscos e (iv) Comparação entre as técnicas.

- **Estudo de avaliação de riscos** – Esta etapa está dividida em 3 atividades: Estado da Arte, Caracterização de projetos e Coleta de Dados.
 - **Estado da Arte** – Esta atividade consiste no estudo do estado da arte da gerencia de projetos com foco no gerenciamento de riscos. Com base neste estudo destacar as técnicas da identificação e avaliação de riscos;
 - **Caracterização de projetos** – Estudo com a finalidade de levantar aspectos fundamentais na ocorrência dos riscos em projetos;
 - **Coleta de Dados** – Elaboração de questionário com base na caracterização dos projetos. Disponibilização e divulgação de

questionário na web a fim de coletar os dados de projetos anteriores com os profissionais de TI.

- **Estudo de técnicas de computação inteligente na avaliação de riscos** – Visa o levantamento das principais técnicas de computação inteligente com potencial para utilização na avaliação de riscos;
- **Estudo e aplicação de redes MLP na avaliação de riscos** – Visa o estudo das RNAs *Multilayer Perceptron* para definição da topologia mais adequada para os dados disponíveis. Treinamento e teste da rede buscando mensurar sua eficiência;
- **Comparação entre as técnicas** - Comparação entre os resultados obtidos e as principais técnicas utilizadas na avaliação de riscos.

1.4 Estrutura do documento

Além deste capítulo introdutório, que apresentou uma visão geral deste trabalho, este documento foi estruturado da seguinte forma:

Capítulo 2 – **Gerenciamento de Riscos em Projeto de Desenvolvimento de Software** – Este capítulo apresenta uma revisão bibliográfica do gerenciamento de riscos em projetos de software. Destaca as técnicas de avaliação de riscos mais utilizadas.

Capítulo 3 – **Computação Inteligente** – Este capítulo apresenta uma revisão bibliográfica das principais técnicas de computação inteligente com potencial para avaliação de riscos: Árvores de Decisão, RBC, Lógica Fuzzy e Redes Bayesianas. Apresenta as Redes Neurais Artificiais com destaque para as Redes MLP.

Capítulo 4 – **Avaliação de Riscos com base em MLP** – Este capítulo apresenta como se deu a definição da caracterização dos projetos e conseqüentemente a coleta de dados. Apresenta a topologia da rede definida e o desempenho da mesma. Realiza uma avaliação dos resultados obtidos pela rede MLP comparando com as demais técnicas mostradas no capítulo 3.

Capítulo 5 – **Conclusão e Trabalhos Futuros** – Este capítulo tem a finalidade de apresentar os resultados obtidos por este trabalho. Além disto apresenta os trabalhos que se relacionam com este e as possibilidades de estudos posteriores a fim de qualificar a técnica proposta.

Capítulo 2

Gerenciamento de Riscos em Projeto de Desenvolvimento de Software

Este capítulo tem o objetivo de apresentar uma visão geral à cerca dos conceitos e práticas no gerenciamento de risco em projetos de desenvolvimento de software.

Este capítulo se torna importante, pois organizar o trabalho na forma de projetos é uma prática bastante difundida entre as organizações ou mesmo na academia. Contudo a base teórica que sustenta esta forma de organização do trabalho é muitas vezes negligenciada. Com o intuito de embasar melhor este estudo será feita uma revisão na literatura da Gerência de Projetos para definir bem os conceitos aqui utilizados.

2.1 Gerenciamento de Projetos

A ampla utilização do termo projeto tem gerado um fenômeno de distorção de seu significado original. Atualmente a palavra projeto tem sido utilizada para qualquer atividade humana, em alguns casos sendo utilizada como sinônimo de trabalho.

O conceito projeto está sim relacionado ao esforço humano, mas não a qualquer esforço. Projetos são, antes de tudo, abstrações de esforço de organização do trabalho humano na direção de um ou mais objetivos. Podem ser de curta ou longa duração, mas têm como característica o fato de transcorrerem em um espaço de tempo determinado. Outras peculiaridades dos projetos são a geração de um produto ou serviço único e a sua elaboração progressiva dividida em etapas [17].

Com a caracterização do que é projeto fica mais clara a separação com a produção ou prestação de um serviço constante. A diferenciação é importante, pois por suas características finitas os projetos necessitam de uma atenção especial a alguns fatores como: tempo, escopo e comunicação. Isto não significa que os processos contínuos não necessitem de atenção a estes pontos, apenas são diferentes, e por isto devem ter suas particularidades respeitadas.

Esta divisão se torna tênue no caso de uma forma de gestão de produção conhecida como Gestão por Projetos, que consiste na aplicação dos conhecimentos e práticas da disciplina de gestão de projetos nos processos contínuos. Nestes casos os processos contínuos são encarados como projetos [15].

A Gestão por Projetos se mostra eficiente na execução de processos contínuos, pois ao se apropriar das premissas de projetos obtém vantagens a partir da definição de metas, prazos e custos mais rígidos. Porém o contrário não se mostra eficiente, tratar um projeto como um processo contínuo pode trazer conseqüências desastrosas. Mesmo assim, inconscientemente muitos projetos são levados pelos seus colaboradores por este caminho.

2.1.1 Ciclo de vida de projeto

Com a finalidade de melhorar o acompanhamento e controle, os projetos são divididos em fases, a composição destas fases compreende o ciclo de vida de um projeto. Nesta divisão são estabelecidas as técnicas de trabalho e pessoal envolvidos em cada fase.

Dentro das fases são definidos marcos (milestones) que são pontos de referência para verificar e medir a execução das tarefas dentro do projeto. Já a conclusão das fases caracteriza o momento de revisão dos trabalhos para detectar os possíveis desvios e efetuar a correção da rota.

2.1.2 Guia PMBOK

O PMI (Project Management Institute) é uma instituição que existe desde 1969, e foi formado a partir da associação de profissionais da área de gerenciamento de projetos. Em 1986 o PMI efetuou a primeira sistematização dos

conhecimentos e melhores práticas em gerência de projetos no documento intitulado: PMBOK (Project Management Body of Knowledge).

O Guia PMBOK organiza os processos de gerenciamento em cinco grupos: iniciação, planejamento, execução, controle e finalização [17]. Além disto, os processos estão organizados por áreas do conhecimento, sendo elas: Gerenciamento da integração do projeto, Gerenciamento do escopo do projeto, Gerenciamento do prazo do projeto, Gerenciamento do custo do projeto, Gerenciamento da qualidade do projeto, Gerenciamento dos recursos humanos do projeto, Gerenciamento da comunicação do projeto, **Gerenciamento dos riscos do projeto** e Gerenciamento das aquisições do projeto.

Os processos de gerenciamento são executados de forma paralela durante o ciclo de vida do projeto, porém o esforço ou intensidade do trabalho varia com sua evolução. A Figura 2.1 apresenta um gráfico com distribuição do esforço durante o ciclo de vida de um projeto.

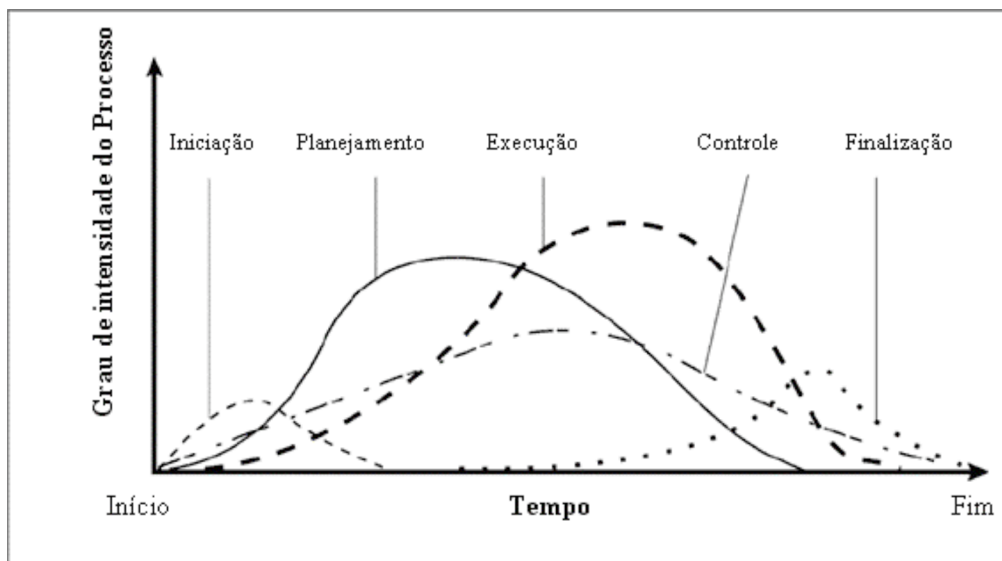


Figura 2.1. Gráfico da intensidade do processo versus tempo proposto no PMBOK [17].

2.1.3 CMMI (SEI)

Em 1987 o SEI (Software Engineering Institute), entidade vinculada a Universidade Carnegie Mellon, lançou sua sistematização do conhecimento a cerca

da maturidade de processos CMM (Capability Maturity Model), posteriormente após uma evolução foi desenvolvido o SW-CMM (Capability Maturity Model for Software).

O CMMI (Capability Maturity Model Integration) é uma evolução do SW-CMM, que visa oferecer uma avaliação e melhoria de processos organizacionais de forma efetiva e eficiente, reduz os custos de formação e avaliação; promove uma visão integrada da melhoria dos processos organizacionais; e um novo meio de representação da informação de disciplinas específicas, através do uso de modelos de melhoria testados [20].

O CMMI propõe cinco níveis de maturidade, quanto maior o nível de maturidade de uma empresa maior a produtividade e qualidade, conseqüentemente menor o risco. A Gerência de Riscos pode dar-se a partir do nível 2, sendo destacada no nível 3, conforme Tabela 2.1.

Tabela 2.1. Áreas de Processos no CMMI

NÍVEL	FOCO	ÁREA DE PROCESSO
1 - Inicial	Não avaliado	Não avaliado
2 - Gerenciado	Gerenciamento básico de projetos	Gerenciamento de requisitos Planejamento de projeto Controle e monitoração de projeto Gerenciamento de suprimentos Avaliação e análise Garantia da qualidade do processo e produto Configuração e gerenciamento
3 - Definido	Padronização de processos	Requisitos de desenvolvimento Soluções técnicas Integração de produtos Verificação Validação Foco no processo organizacional Definição do processo organizacional Treinamento organizacional Gerenciamento de projeto integrado Gerenciamento de riscos Integração da equipe de trabalho Gerenciamento integrado de suprimentos Análise de decisões Ambiente organizacional para integração
4 - Gerenciado	Gerenciamento	Performance organizacional do

quantitativamente	quantitativo	processo Gerenciamento quantitativo de projetos
5 - Otimizado	Melhoramento contínuo de processo	Inovação organizacional Análise de causas e resoluções

Os processo são divididos em quatro áreas: Gerência de processos, Gerência de projetos, Engenharia e Apoio. A Gerência de Riscos está situada dentro dos processos de gerência de projetos.

2.1.4 Normas ISO 9000-3, ISO 12207 e ISO 15504

A norma ISO 9000-3 é um guia de aplicação da norma ISO 9001 para desenvolvimento, fornecimento e manutenção de software. A norma ISO 9001 faz parte da série 9000 voltada para gestão e garantia da qualidade. As diretrizes propostas na norma ISO 9000-3 cobrem questões como: a garantia do entendimento comum entre as partes à cerca dos requisitos, a utilização de metodologias consistentes para desenvolvimento de software e gerenciamento do projeto como um todo. Com relação à Gerência de Riscos esta norma não aborda explicitamente, mas, apresenta as atividades de identificação, análise, controle e monitoração de riscos.

A norma ISO 12207 é a primeira norma internacional que descreve em detalhes os processos, atividades e tarefas que envolvem o fornecimento, desenvolvimento, operação e manutenção de produtos de software. Sua principal finalidade é servir de referência para os demais padrões que venha a surgir.

O projeto SPICE (Software Process Improvement and Capability dEtermination), baseado em modelos como o CMM, resultou na elaboração da norma ISO 15504. O processo Gerência de Riscos é definido nesta norma.

2.2 Gerenciamento de riscos

Nesta seção abordaremos os aspectos mais relevantes da disciplina de Gerência de Riscos. A Seção 2.1 deste trabalho apresentou uma visão geral dos

conhecimentos e práticas mais avançadas utilizadas pelas organizações para gerenciamento de projetos de desenvolvimento de software. Nestes modelos, normas ou guias compilados a disciplina de Gerência de Riscos é abordada.

Não é à toa que os documentos que compilam as boas práticas abordem este tema. A entrega de software com qualidade e respeitando prazo e orçamento estipulado que atenda as necessidades do usuário sempre foi um sonho buscado. Mesmo com todas as metodologias, sistemas de suporte ao gerente e com a aprendizagem acumulada por parte das organizações, os números não são muito animadores. O Standish Group, em 1994 publicou o Chaos Report [22] mostrando o resultado de uma pesquisa realizada com mais de 175.000 projetos de software no mundo. Esta pesquisa concluiu que apenas 16% dos projetos são concluídos com sucesso.

O resultado desta pesquisa concluída em 1994 é estarrecedor, pois 25 anos após a crise do software ela parece não acabar. Reforçando ainda mais esta tese os relatórios seguintes apresentaram resultados semelhantes, como pode ser visualizado na Figura 2.2.

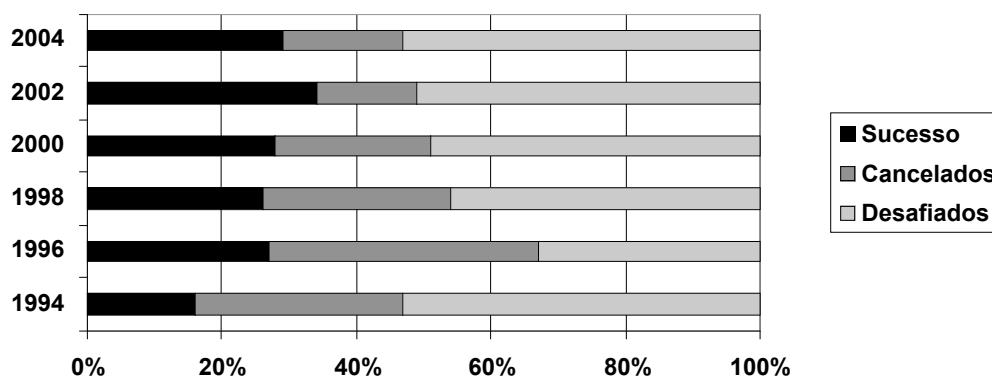


Figura 2.2. Gráfico com resultado Chaos Report [22]

Estes dados são muito importantes por explicitar o problema e promover um debate à cerca dos motivos que tem levado a uma evolução tão lenta no índice de sucesso em projetos de software.

Um aspecto deve ser colocado em evidência, pois se apenas um quarto dos projetos ao seu término foram considerados bem sucedidos, então os riscos em empreender nesta área são grandes. Por outro lado, poucos estudos que apresentem alternativas são realizados nesta área e menor ainda é a oferta de ferramentas de suporte à gestão de riscos em projetos que atendam as necessidades dos gerentes de projetos. Com estes elementos fica fácil compreender a atual situação, pois as possibilidades de encontrar problemas no caminho são grandes, mas as organizações não estão preparadas para identificá-los e encontrar alternativas em tempo de reverter à situação.

Ter a possibilidade de se antecipar aos problemas para tomar decisões certas e rápidas passa cada vez mais a ser fundamental para todas as organizações. Com o mercado cada vez mais competitivo e exigente isto passa a ser essencial para sobrevivência no mercado. As pequenas organizações que dispõem de orçamento curto e em geral alocam sua equipe em mais de um projeto necessitam dar este salto de qualidade incorporando os processos de Gerenciamento de Risco em seus projetos, do contrário será muito difícil à sobrevivência em um mercado onde menos de um terço dos clientes ficam satisfeitos.

Um dos fatores que contribuem para o agravamento da questão reside na incompreensão à cerca deste tema pelos gerentes dos projetos. O próprio conceito de risco é visto por diferentes modos dentro das organizações. Para um tratamento adequado é de fundamental importância que exista um consenso entre os membros de um projeto da forma que a instituição encara os riscos. Outro aspecto relevante é o entendimento que não basta identificar os fatores de riscos é fundamental que se faça uma gestão destes riscos identificados.

Embora não exista consenso neste ponto, tradicionalmente a Gerência de Riscos é vista como uma parte da Gerência de Projetos. Na literatura podemos encontrar pontos de vistas diferentes, mas, independente disto há a concordância da importância de um gerenciamento proativo, estruturado e consistente dos riscos para o sucesso dos projetos de uma organização.

Dentre as sistematizações discutidas na Seção 2.1 utilizaremos o PMBOK para uma visão geral do processo de gerenciamento de riscos.

2.2.1 Visão do PMBOK

O Guia PMBOK organiza o processo de Gerência de Riscos em seis processos, sendo eles:

- **Planejamento do gerenciamento de riscos** – Fase em que se definirá como serão realizadas as atividades de gerenciamento de risco do projeto. Esta fase é fundamental, pois define as tarefas a partir das características dos projetos.
- **Identificação de Riscos** – Nesta fase serão determinados quais os riscos que deverão permear o projeto. Esta fase será melhor abordada na Seção 2.3.
- **Análise qualitativa de riscos** – Fase com intuito de priorizar os risco com base na probabilidade de ocorrência e o impacto dos riscos identificados.
- **Análise quantitativa de riscos** – Nesta fase serão mensurados a probabilidade de ocorrência e o impacto da incidência dos riscos sobre o projeto.
- **Planejamento de resposta a riscos** – Elaboração de plano de reação aos riscos avaliados durante a execução do projeto.
- **Monitoramento e controle de riscos** – É o momento de eliminar também, os riscos que não vão afetar mais o projeto, e de monitorar, para verificar se as estratégias de mitigação e prevenção estão obtendo sucesso.

2.3 Técnicas de avaliação de riscos

Nesta seção serão apresentados os métodos mais utilizados para identificação e análise de riscos. É importante observar as particularidades de cada etapa. Embora as atividades possam se confundir, ambas tem objetivos distintos porém complementares.

A identificação dos riscos é feita com o intuito de levantar quais os riscos que devem ser considerados no projeto e relatar suas características a fim de possibilitar a sua gestão.

Já a análise dos riscos é uma etapa posterior que consiste no estudo dos riscos identificados com o objetivo de definir a possibilidade da ocorrência deste e de avaliar o impacto gerado no projeto. Estas informações são subsídios para definir a estratégia de mitigação e prevenção dos riscos e o plano de contingência.

2.3.1 Métodos de Identificação de Riscos

Diversos métodos de identificação são recomendados pela literatura de Gerência de Riscos. Os principais métodos aplicados são: tempestade de idéias (*brainstorming*), listas de verificação (*checklist*), comparação por analogia, análise de premissas, decomposição, técnicas de diagramação, técnica delphi, revisão de documentação e entrevistas.

Em pesquisa realizada pelo PMI as preferências se concentram no brainstorming e entrevista conforme gráfico apresentado na Figura 2.3.

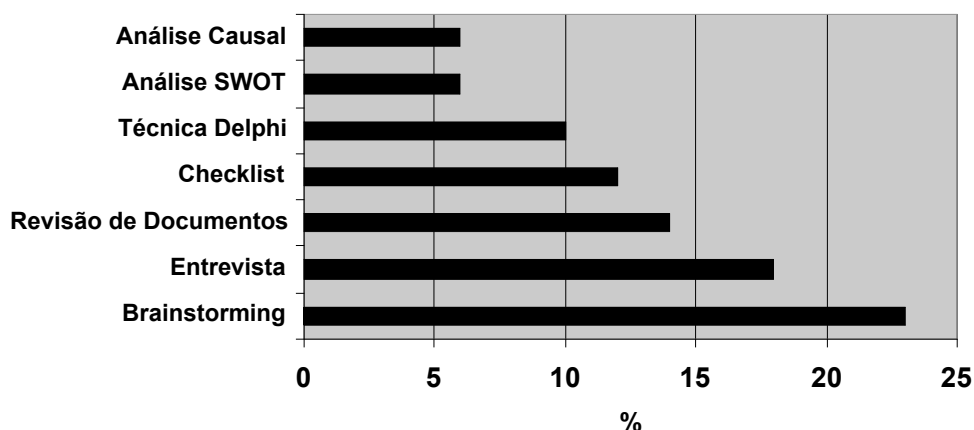


Figura 2.3. Gráfico com utilização das técnicas de identificação [9].

A tabela apresenta uma visão geral sobre as principais técnicas de identificação de riscos:

Tabela 2.2. Descrição das técnicas de identificação de riscos

TÉCNICA	DESCRIÇÃO
Tempestade de idéias (<i>Brainstorming</i>)	Esta técnica consiste na reunião com especialistas onde as idéias são estimuladas e nenhuma é descartada. Nesta técnica as idéias devem ser registradas sem um julgamento até atingir uma solução considerada adequada.
Entrevista com especialista	Este método tem a finalidade de coletar dados a partir da experiência acumulada pelos especialistas. A utilização de questionários é bastante difundida, podendo ser aplicado individualmente ou através de grupos focais. A vantagem desta técnica é abranger diversas experiências. Uma limitação é o próprio questionário que não deve restringir as respostas dos especialistas.
Revisão de documentos	Esta é uma técnica subjetiva que é realizada com apoio de uma lista de verificação e categorias pré-definidas de riscos. O revisor lê o documento e tenta verificar não conformidades. Esta técnica é muito utilizada para identificação de riscos técnicos.
Lista de verificação (<i>checklist</i>)	Método onde diversos aspectos são analisados a partir de uma lista com itens pré-definidos. Estes itens são definidos em projetos similares. Uma deficiência desta técnica está na impossibilidade de listar todos os riscos. Em geral a utilização do checklist vêm para reforçar outra técnica utilizada, como <i>brainstorming</i> .
Técnica Delphi	Variante dos grupos focais esta técnica é utilizada a fim de obter consenso entre grupos de especialistas. O consenso é obtido a partir de rodadas de comentários a partir de um tema específico com ajuda de um facilitador. Têm a vantagem de atingir um equilíbrio entre as influências dos especialistas, mas está limitado ao tema proposto no questionário, limitando-se a troca de informações.
Análise SWOT	A análise SWOT (Strength, Weakness, Opportunity and Threats) foca na análise da organização e não de um projeto específico [17]. Os pontos fracos e fortes levantados na organização são fontes de informação para levantamento de riscos em um projeto específico da organização.
Análise Causal	É baseado na relação causa/efeito com o intuito de identificar a origem dos riscos. Entre os métodos que utilizam a Análise Causal estão diagrama de Causa e Efeito (Espinha de Peixe ou Fishbone) e a técnica do 6 W's (Who, Why, What, Whichway, Wherewithal, Whe) [2].
Comparação por analogia	Esta técnica consiste na comparação entre projetos a partir das suas similaridades. É uma técnica simples, mas incute uma subjetividade no processo a partir da interpretação dos dados históricos.
Análise de Premissas	Uma possibilidade de risco que vai além dos riscos inerentes a execução do projeto são riscos decorrentes das hipóteses premissas adotadas no planejamento. Nesta técnica as

	premissas são identificadas e validadas durante a execução do projeto.
--	--

2.3.2 Análise Qualitativa de Riscos

A Análise qualitativa de riscos é a forma mais utilizada de análise de riscos. Para isto utiliza a experiência dos especialistas com o intuito de medir o impacto da incidência dos riscos no projeto.

A principal técnica de avaliação qualitativa descrita pelo Guia PMBOK [17] é a matriz de riscos. Esta técnica consiste em estimar a probabilidade de ocorrência e a gravidade do risco por meio de uma escala e a partir destes dados efetuar o cálculo de exposição ao risco. Este dado pode ser verificado pela Equação 2.1.

Equação 2.1. Cálculo da Exposição ao Risco.

$$E(n) = P(n) \times I(n)$$

$E(n)$ = Exposição ao risco n

$P(n)$ = Probabilidade de ocorrência do risco n

$I(n)$ = Impacto do risco n

Os valores utilizados para realizar o cálculo de exposição ao risco são definidos a partir de alguma escala, onde os conceitos como alto, médio e baixo são convertidos em valores numéricos.

2.3.3 Técnicas Quantitativas de Análise de Riscos

A necessidade em medir é fundamental para o monitoramento de um projeto. A partir de indicadores os planos de contingência podem ser iniciados favorecendo assim uma reação em tempo hábil.

A utilização de métricas para determinar a possibilidade de ocorrência de determinado risco ou mesmo o impacto provocado caso este problema se consolide são os pilares de um monitoramento efetivo dos riscos em um projeto.

Em contraposição a análise qualitativa, que prima pelo entendimento à cerca dos riscos, a análise quantitativa foca em medir o risco, e a partir desta característica é que se observa seu maior potencial. As análises quantitativas podem ser realizadas a qualquer tempo, uma vez que não necessita reunir grupos de especialistas a cada rodada do projeto. As alternativas mais utilizadas são descritas na Tabela 2.3.

Tabela 2.3. Descrição das técnicas de análise qualitativa de riscos

TÉCNICA	DESCRIÇÃO
Árvores de Decisão	Técnica de Inteligência Artificial que divide o problema em subconjuntos formando a árvore de decisão, onde cada subconjunto do problema é um nó. As decisões podem ser tomadas a partir do cálculo da entropia ou do cálculo de exposição ao risco. Esta técnica e o cálculo de entropia serão abordados no Capítulo 3 e o cálculo de exposição ao risco será detalhado na subseção 2.3.3.
Técnicas de Simulação	Estas técnicas são realizadas a partir de software específico, a análise mais comum é a Monte Carlo. A partir de critérios como quantidade de horas e custo de execução são feitas projeções sobre o projeto em avaliação [9].

A utilização das técnicas quantitativa de avaliação de riscos tende a obter resultados melhores com o tempo uma vez que o registro dos históricos dos projetos passados favorecem a avaliação futura.

2.4 Resumo do Capítulo

Este capítulo apresentou uma visão geral do gerenciamento de projeto destacando a importância observada a partir da literatura de referência (Guia PMBOK, CMMI, ISO 9000-3, ISO 12207 e ISO 15504) do gerenciamento de riscos.

No gerenciamento de riscos a visão do Guia PMBOK foi apresentada com o intuito de fornecer um panorama completo das atividades sugeridas pelo PMI. Isto proporciona uma visão global do processo favorecendo um melhor entendimento a cerca da avaliação de riscos.

Nos aspectos referentes à avaliação dos riscos o objetivo foi reunir as principais técnicas de identificação, análises qualitativas e quantitativas de riscos.

Esta revisão bibliográfica foi realizada com o objetivo de mostrar o estado da arte atual em gerenciamento de projetos e riscos. Pois, são pré-requisitos fundamentais para compreensão da importância do estudo de novas técnicas de avaliação de riscos.

Capítulo 3

Computação Inteligente

A Inteligência Artificial (IA) é uma área da Ciência da Computação que estuda o desenvolvimento de técnicas computacionais inspiradas na capacidade humana de resolver problemas e sobretudo aprender com eles. Entre os estudiosos da área existe um debate que a divide entre IA fraca e IA forte. No cerne desta questão reside no fato de ser possível ou não construir uma máquina consciente. Como o objetivo deste trabalho é apenas a aplicação de uma técnica consagrada de IA foi feita a opção por utilizar uma nomenclatura mais adequada para este fim que é a denominação deste conjunto do conhecimento de computação inteligente.

Neste capítulo serão apresentadas as técnicas de computação inteligente mais utilizadas e as com potencial de aplicação na avaliação de riscos de projetos, com o intuito de fornecer uma visão geral à cerca da aplicação destas técnicas no domínio da avaliação de riscos.

3.1 Técnicas de computação inteligente

As técnicas de computação inteligente tiveram seu desenvolvimento inicial logo após a segunda guerra mundial com a publicação do artigo de Alan Turing intitulado “Computing Machinery and Intelligence” [24]. Inspirados na capacidade de aprender do homem foram desenvolvidos alguns modelos computacionais. Porém só com o surgimento do computador moderno estas técnicas se tornaram aplicáveis na solução de problemas reais.

Atualmente a computação inteligente é utilizada para uma gama de aplicações que vão desde sistemas de reconhecimento de voz ao *Deep Blue* da IBM (programa de computador que derrotou o campeão mundial de xadrez Garry Kasparov).

Como as técnicas de computação inteligente apresentam excelentes resultados na tarefa de classificação, são portanto candidatos naturais para auxiliar nos desafios encontrados no gerenciamento de riscos.

3.1.1 Árvores de Decisão

Esta técnica de computação inteligente consiste em representar as informações pertinentes a um problema em uma estrutura de árvore. Na árvore de decisão os nós internos são testes a serem realizados e os nós folha são as decisões como mostrado na Figura 3.1.

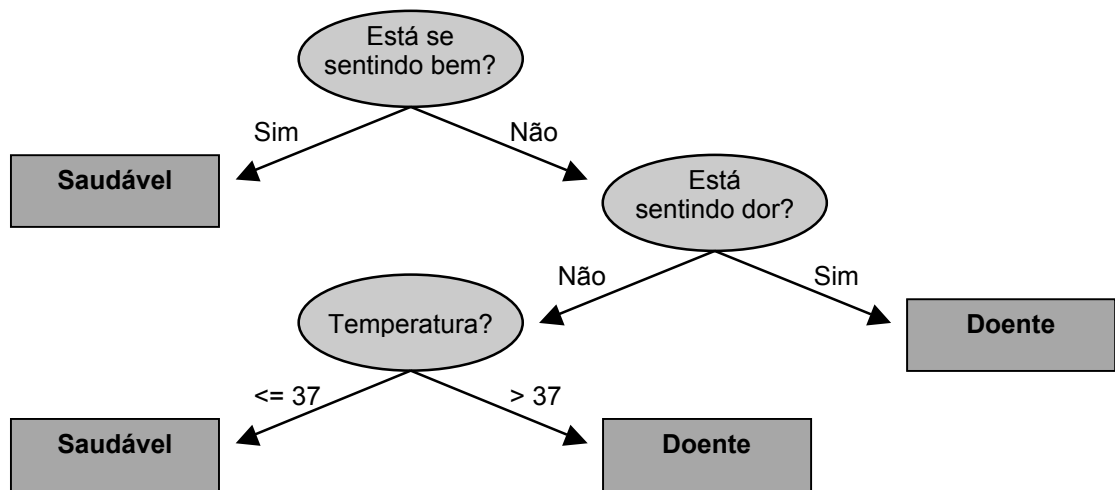


Figura 3.1. Ilustração de uma árvore de decisão hipotética [16]

Para a construção de uma árvore de decisão adequada a posição do atributo que será testado é bastante relevante, pois influencia diretamente no tamanho final da árvore [16]. Desta forma, os teste mais significativos para a solução do problema devem ser posicionados nos níveis mais baixos. A seleção dos atributos pode ser realizada a partir de métricas, a mais utilizada entre elas é a função de entropia.

Após a utilização da função de entropia para determinar o grau de relevância dos atributos pode ser feito um processo de melhoria de desempenho a partir da inclusão de novos ramos ou simplesmente pelo balanceamento da árvore. A partir da árvore resultante o processo de utilização consiste em percorrer a árvore do nó raiz até encontrar uma folha que consiste no resultado encontrado.

A definição na árvore de decisão é uma etapa crucial e alguns cuidados devem ser considerados. Existe a possibilidade da árvore gerada seja muito específica, perdendo assim sua capacidade de generalização. Esse problema é conhecido como overfitting e pode ser resolvido realizando um processo chamado de poda [16]. Como o próprio nome remete, a poda consiste num processo de redução da árvore retirando ramos pouco relevantes para a solução do problema.

3.1.2 Raciocínio Baseado em Casos (RBC)

O Raciocínio baseado em casos (RBC) como o próprio nome sugere consiste numa técnica de computação inteligente que utiliza os dados de projetos passados na solução de problemas futuros [11].

A base do RBC consiste na busca por problemas anteriores que se aproximem do problema a ser resolvido. Caso os problemas forem suficientemente semelhantes à solução de um problema pode ser aplicada no outro.

As comparações são feitas a partir de uma técnica conhecida como cálculo de similaridade entre casos. Como fica fácil perceber a definição de similaridade é o aspecto mais relevante a ser considerado na compreensão do sistema [11]. A busca da similaridade pode dar-se utilizando as seguintes propriedades:

- Reflexividade: Um caso é similar a si mesmo;
- Simetria: Se uma caso é similar ao outro a recíproca é verdadeira;
- Transitividade: Se um é similar ao outro e este é similar com um terceiro, então o primeiro caso é similar ao terceiro.

Para a resolução dos problemas baseando-se nas experiências passadas a etapa da recuperação dos dados relevantes ao problema atual pode ser dividida em três etapas. A primeira é o assessoramento da situação, que consiste basicamente no fornecimento das informações à cerca do problema por parte do usuário para o sistema. A segunda etapa é o casamento, que consiste em uma consulta no banco de dados e a comparação entre os casos encontrados com o caso atual. A conclusão da recuperação se dá através da seleção, onde os casos mais

semelhantes são escolhidos a partir do cálculo da similaridade de casos de forma detalhada considerando o domínio de aplicação dos dados [11].

3.1.3 Lógica Fuzzy (Difusa)

A capacidade de retirar conclusão de informações imprecisas é uma discussão que remonta o século V a.C., onde no poema de Parmênides de Eléia o autor separa as coisas entre o ser e não ser. Este debate filosófico percorreu os séculos passando por Aristóteles, Frege, Bertrand Russell e Kant. Apenas em 1920 os primeiros passos para construção de uma lógica mais flexível foram dados pelo Polonês Jan Lukasiewicz, que formulou uma lógica com 3 valores (0, 1/2 e 1). Só em 1965 em um artigo de Lofti Asker Zadeh sistematizou o problema dos conceitos vagos e chamou de teoria dos conjuntos fuzzy ou nebulosos. A partir da década de 70 é que os primeiros grupos de pesquisa nesta área foram formados. Em 1976 surgiu a primeira aplicação industrial de Lógica Fuzzy com objetivo de controlar fornos de cimento. A partir disto vários sistemas de controle utilizando fuzzy foram desenvolvidos [21].

A Lógica Fuzzy ou Lógica Difusa pode ser vista como uma extensão da Lógica Booleana. Na lógica Booleana os valores são bem definidos, ou como chamam os estudiosos de Fuzzy, são valores nítidos. Isto significa que uma variável assume valores que podem ser 0 ou 1 (falso ou verdadeiro). Já na Lógica Fuzzy são possíveis valores intermediários entre 0 e 1. Neste caso a variável assume um grau de pertinência a um determinado grupo. A diferença entre as duas lógicas é ilustrada na Figura 3.2 [7].

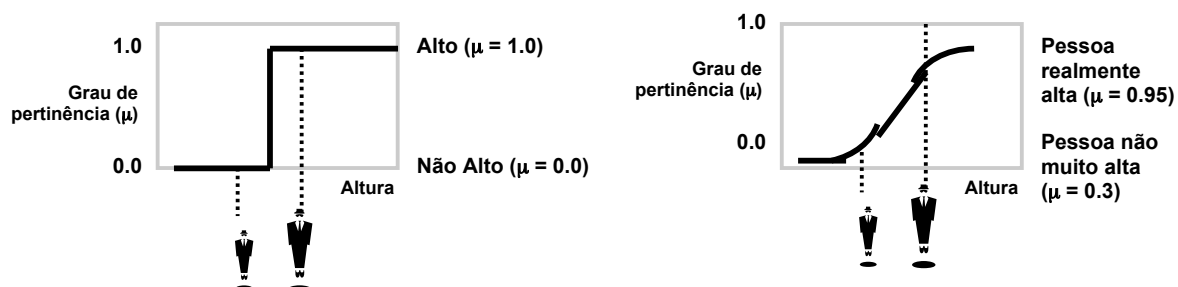


Figura 3.2. Lógica Booleana x Lógica Fuzzy

Assim a Lógica Fuzzy se apresenta como uma alternativa flexível que possibilita a aplicação em diversos problemas do mundo real, como os sistemas de suporte a tomada de decisão, em que as opções não são tão precisas. Como métodos de classificação que são baseados na Lógica Fuzzy podemos destacar: Fuzzy C-Means, Fuzzy Nearest Neighbours, método baseado em Regras Fuzzy e Fuzzy Pattern Matching method [7].

3.1.4 Redes Bayesianas

Redes Bayesianas é uma técnica para tomada de decisão baseada em probabilidade condicional. A partir da década de 90 as pesquisas e a aplicação desta técnica se intensificaram. Desde então as Redes Bayesianas têm sido utilizadas para resolver problemas em diversas áreas.

O conceito de probabilidade condicional é fundamental para a compreensão do Teorema de Bayes. A probabilidade condicional $P(A|B)$ difere da probabilidade simples. Neste caso devemos considerar a probabilidade do evento A ocorrer dado que o evento B ocorreu. Em outros termos temos: “probabilidade de A dado B”.

O teorema de Bayes se utiliza também do teorema das Probabilidades Totais. Este teorema mostra que para qualquer evento B podemos obter a probabilidade de sua ocorrência como a combinação do espaço amostral. Com esta base o teorema de Bayes é enunciado conforme a equação 3.1.

Equação 3.1. Teorema de Bayes

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

Com base no teorema enunciado as Redes Bayesianas são construídas a partir de grafos que ilustram as relações de causalidades entre as variáveis. Nestes grafos as variáveis são os nós dos grafos e os relacionamentos entre as variáveis são os arcos dos grafos [8].

Para ilustrar a construção de uma Rede Bayesiana vamos utilizar o clássico caso de João, Maria e o alarme contra ladrões. Neste caso um alarme foi instalado,

este alarme é razoavelmente confiável na detecção de assaltos, mas também responde ocasionalmente a pequenos terremotos. Ao sair de casa você pede a seus vizinhos João e Maria para ligarem quando ouvirem o alarme. João telefona sempre que houve o alarme, mas às vezes confunde com o toque do telefone e liga. Já Maria pelo fato de ouvir música muito alta por vezes não ouve o alarme. Com este cenário podemos estimar a seguinte Tabela 3.1 de probabilidades:

Tabela 3.1. Probabilidade de o alarme tocar [8]

Entrada 1	Entrada 2	P(A L,T)	
		V	F
Ladrão	Terremoto	0,95	0,05
Ladrão	Não Terremoto	0,95	0,05
Não Ladrão	Terremoto	0,29	0,71
Não Ladrão	Não Terremoto	0,001	0,999

Neste caso se quisermos calcular a probabilidade de João ligar caso tenha um ladrão na casa e não houve um terremoto, então temos:

$$P(J) = P(J|A) \cdot P(A|R \cap \neg T) = 0,95 \cdot 0,9 = 0,855 \text{ ou } 85,5\%$$

Com base nesta idéia a Rede pode ser construída como na ilustração da Figura 3.3.

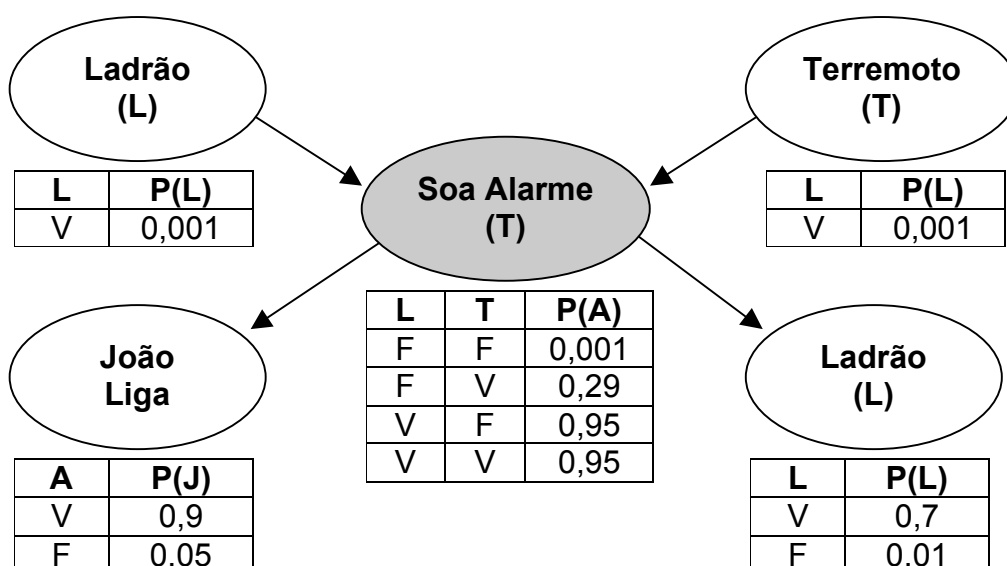


Figura 3.3. Topologia com a probabilidade do alarme soa da Rede Bayesiana [8].

Na figura 3.3 pode ser visto a Rede Bayesiana construída para o problema em questão. É facilmente percebido o grafo resultante com as probabilidades calculadas.

3.2 Redes neurais artificiais

As Redes Neurais Artificiais (RNA) são modelos matemáticos que se assemelham às estruturas neuronais biológicas e que têm capacidade computacional adquirida por meio de aprendizado e generalização [3]. Inspirados pelo sistema neurológico as RNAs são constituídas por unidades de processamento simples e distribuídas chamadas de neurônio assim como no modelo biológico. A força da conexão entre os neurônios é conhecida como pesos sinápticos e armazenam o conhecimento adquirido.

As RNAs são relativamente recentes onde a pouco mais de meio século foi apresentado o trabalho pioneiro dos neurofisiologistas Donald Hebb e Karl Lashley na década de 40 [12]. Neste trabalho as simulações foram feitas a base de papel e lápis. Em 1943 McCulloch e Pitts propuseram no clássico artigo *A logical calculus of the ideas immanent in nervous activity* onde propuseram um modelo simplificado de neurônio como pode ser visto na Figura 3.4.

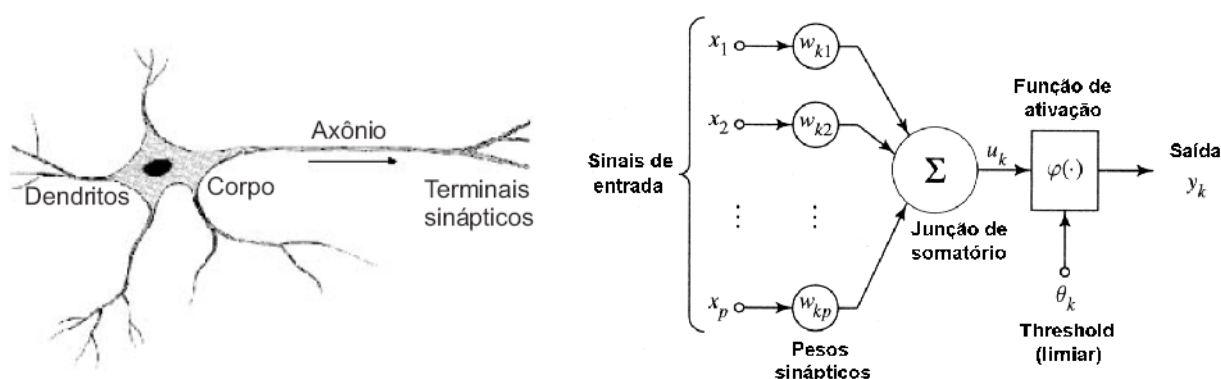


Figura 3.4. Modelo simplificado de neurônio biológico e proposto por McCulloch e Pitts.

O neurônio biológico pode ter seu funcionamento simplificado para facilitar a compreensão. Os dendritos recebem os “sinais” de entrada do neurônio, estes sinais

são decorrentes de estímulos enviados por outros neurônios. Os “sinais” influenciam de forma diferenciada cada neurônio. O grau de influências de um neurônio sobre outro se modifica com o tempo em decorrência principalmente da aprendizagem. Os estímulos recebidos pelo corpo do neurônio são “somados” e caso superem um certo limiar de ativação este neurônio “dispara” um estímulo para outros neurônios ligados a este.

O neurônio artificial proposto por McCulloch e Pitts tem um funcionamento muito semelhante ao modelo biológico sendo composto por Sinais de entrada, Pesos Sinápticos, uma Junção de somatório, uma função de ativação (ou Threshold) e a saída que se comportam como terminais sinápticos, dendritos, corpo e axônio.

Os sinais de entrada são as variáveis que serão utilizadas na RNA. Os pesos sinápticos são valores que são ajustados durante a aprendizagem da rede. Estes valores são multiplicados pelos valores de entrada e determinam a influência do sinal de entrada no neurônio. Os pesos podem assumir valores positivos e negativos, os valores positivos podem ser associados a sinapses excitatórias e no caso de valores negativos à sinapse inibitória dos neurônios biológicos. A função de junção somatório tem a finalidade de somar a contribuição de cada sinal de entrada multiplicada pelo seu peso gerando o valor que será passado como parâmetro para função de ativação. A função de ativação consiste numa função que define um limiar de ativação, ou seja, define se o neurônio será ativado. Várias funções podem ser utilizadas para este fim tais como: rampa, degrau, sigmóide (ou logística) e hiperbólica.

Este modelo desenvolvido na década 40 foi bastante disseminado e em 1958 o livro de “Principles of neurodynamics de Rossenblatt” sistematizou várias idéias sobre os *Perceptrons*, que são modelos de neurônios baseados no modelo de McCulloch e Pitts [12]. Após este período os algoritmos de Perceptron foram desenvolvidos e aperfeiçoados. Mas no final da década de 60 o estudo de Minsky e Papert provaram que os modelos apresentados até então resolviam apenas problemas de associação de padrões linearmente separáveis. Com esta descoberta as RNAs passaram as duas décadas seguintes relegadas a um plano secundário.

Com o desenvolvimento da computação e principalmente pela contribuição de Rumelhart, Hinton e Willians que resolveram o problema de associação a padrões não-linear com a criação da Regra Delta Generalizada, mais conhecida como Algoritmo de correção de erros de Retropropagação para Redes Perceptron de várias camadas (Multilayer Perceptron). Isto impulsionou o desenvolvimento das RNAs que vem sendo pesquisada e aplicada em diversas áreas desde então [16].

Com desenvolvimento das pesquisas a cerca das RNAs várias formas de organizar os neurônios foram sistematizadas, criando assim o conceito de topologia das Redes Neurais Artificiais. A topologia consiste na determinação das características da rede como: a quantidade de neurônios, a quantidade de camadas (Multilayer Perceptron) ou de forma cíclica (Redes de Hopfield). A topologia ou arquitetura das RNAs apresenta um gama de alternativas que possuem características que favorecem a aplicação em domínio específico. Ou seja, a escolha da RNA mais apropriada depende do problema a ser resolvido.

Além da topologia da RNA é importante definir a forma de aprendizagem da rede. Basicamente existem duas possibilidades o aprendizado supervisionado e o não supervisionado. No método supervisionado um conjunto de entradas e as saídas esperadas são fornecidos para que a rede ajuste os seus pesos (sinápticos). Já no método não supervisionado não existe o supervisor para fornecer os resultados esperados para um conjunto de entradas, assim a rede deve buscar associações relevantes a partir da extração das propriedades estatísticas exclusivamente com os dados de entrada criando classes e grupos representativos.

Na seção seguinte será realizada uma apresentação mais detalhada das Redes Multilayer *Perceptron* (MLP). As MLP são as RNAs mais utilizadas, por obter bons resultados para a maior parte das aplicações.

3.1.5 MLP – Multilayer Perceptron

As Redes MLP são constituídas de no mínimo três camadas de neurônios, sendo uma camada de entrada e uma de saída e no mínimo uma camada intermediária (ou escondida). Os sinais de entrada são propagados de uma camada para a seguinte até atingir a saída da rede conforme figura 3.5. É importante

observar que cada neurônio de uma camada está ligado a todos os neurônios da camada seguinte.

A rede MLPs tem sido utilizada com grande sucesso na solução de problemas com alto grau de não-linearidade. No treinamento supervisionado das MLPs o algoritmo de retro-propagação do erro para ajustar os pesos das conexões entre os neurônios é utilizado. Basicamente este processo é constituído por duas etapas. Na primeira o sinal é propagado da entrada para a saída (feedforward) e posteriormente a retro-propagação do erro (backpropagation) [16].

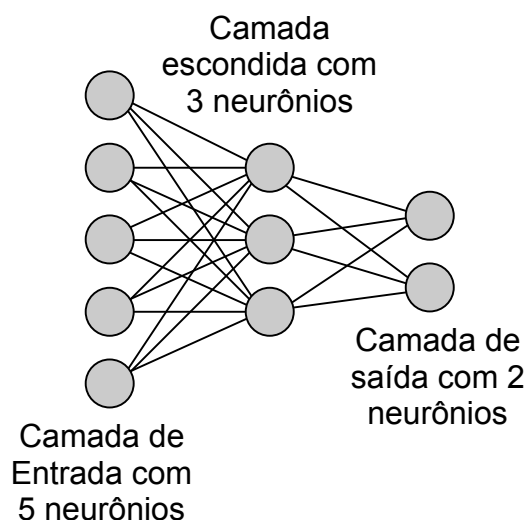


Figura 3.5. RNA Multilayer Perceptron com uma camada escondida

Na fase do *feedforward* o vetor de entrada de dados é apresentado à camada de entrada e estes valores passam pelo processamento dos neurônios desta camada e são propagados para a camada seguinte e este processo se repete a cada camada até atingir a camada de saída. Os sinais nesta etapa são conhecidos como sinal funcional. Nesta etapa os pesos permanecem estáticos.

Na fase do *backpropagation* uma regra de correção de erro é aplicada. Uma das regras utilizadas é a do gradiente descendente que busca minimizar o erro médio quadrático. Após o cálculo do valor a ser ajustado nos pesos o sinal percorre o caminho inverso da rede ajustando o peso dos neurônios. Nesta etapa o este sinal é chamado de sinal de erro.

O processo completo de desenvolvimento de uma rede MLP consiste inicialmente em realizar uma coleta de dados. Como as RNAs trabalham apenas

com valores numéricos é necessário codificar os dados categóricos. Finalmente os dados devem ser divididos para utilizar uma parte no treinamento da rede e outra para teste.

No treinamento de uma rede MLP dois parâmetros são importante para a qualidade do resultado: a taxa de aprendizagem e o momento. A partir deste parametros o treinamento é realizado até atingir os critérios para encerramento do treinamento. A parada no treinamento é feita principalmente a partir da validação cruzada que verifica se o treinamento está produzindo melhorias no resultados. Porém além da validação cruzada o critério do número de épocas é utilizado como limite superior do treinamento. Ou seja, no caso da validação cruzada não parar o treinamento este será finalizado ao atingir o número de épocas estabelecido. É importante observar que a parada do treinamento pelo número de épocas é um indício de que os ajustes não estão produzindo bons resultados ou por subajustamento ou superajustamento.

3.3 Resumo do Capítulo

Este capítulo apresenta as técnicas de computação inteligente que apresentam grande potencial para resolver o problema de identificação e análises de risco, sendo elas: Árvores de Decisão, Raciocínios Baseados em Casos (RBC), Lógica Fuzzy e Redes Bayesianas.

Uma seção específica deste capítulo trata das Redes Neurais Artificiais com destaque para as redes MLP. Esta seção tem o intuito de apresentar os aspectos fundamentais da rede MLP visando compreender a aposta deste estudo como uma técnica viável e eficiente para avaliação de riscos.

Capítulo 4

Avaliação de Riscos com base em MLP

Como visto no capítulo anterior as RNAs consistem em ferramentas poderosas que tem sido aplicada em um gama de aplicações nas mais variadas áreas. A capacidade de identificar padrões e generalizar para outros casos das RNAs e em particular da MLP é o motivo principal da escolha desta técnica para estudo na solução dos problemas pertinentes ao gerenciamento de risco.

Um aspecto na utilização das Redes Neurais Artificiais de deve ser observado. A qualidade dos resultados não está apenas relacionado com a definição dos melhores parametros para treinamento da rede. Outros pontos que merecem atenção especial são: definição dos dados que são relevante na análise do problema, a qualidade dos dados a serem utilizados, as saídas desejadas e a quantidade de informações disponíveis.

Devido ao conjunto reduzido na massa de dados este trabalho focou nos pilares do gerenciamento de risco, desta forma foram implementadas quatro redes MLP para efetuar predição sobre: finalização, prazo, custo e escopo. Os dados referente a retrabalho e sobretrabalho não foram utilizadas neste estudo ficando servindo de base para trabalhos futuros.

4.1 Avaliação baseada em dados passados

Com a finalidade de treinar uma rede MLP para adquirir a capacidade de generalização a partir do histórico de projetos passados, a definição das informações que serão analisadas (dados de entrada e saída) se torna de crucial importância.

Para compreender melhor esta questão podemos pensar de seguinte forma: “para obter boas respostas é fundamental fazer boas perguntas”. Na seção seguinte

será apresentada a opção deste estudo na busca por uma caracterização objetiva e significativa dos projetos.

Não basta apenas fazer a pergunta correta, também é importante saber o que se pretende obter com a resposta. Embora pareça evidente que os profissionais saibam o que estão buscando, nem sempre isto é verdade. É comum a elaboração de indicadores que pouco ajudem na tomada de decisão. Aqui definimos um conjunto bastante restrito de variáveis de saída, conforme apresentado na seção 4.1.2, mas que enfoca no essencial de um projeto.

Um terceiro aspecto que fecha este ciclo com êxito é buscar as respostas das pessoas corretas, fazendo com que as perguntas cheguem a elas. Na seção 4.1.3 será mostrado como se deu a coleta dos dados utilizados neste trabalho.

4.1.1 Caracterização dos projetos (Entrada da Rede)

A caracterização de projetos neste trabalho utilizou a proposta elaborada em Trabalho de Conclusão de Curso [11]. Esta proposta foi baseada na caracterização utilizada no desenvolvimento do Modelo de Adaptação de Processo de Software - MAPS [6]. Para tornar a caracterização mais objetiva apenas um subconjunto da proposta do MAPS foi utilizada. Para enriquecer esta proposta a análise de dois aspectos considerados importantes foram incluídas.

As características eleitas como mais relevantes na proposta do MAPS foram: Tamanho da equipe, Distribuição geográfica da equipe, Experiência da equipe de desenvolvimento e o Tamanho do projeto. Os aspectos não utilizados foram: Padrões adotados, Exigências Contratuais, Ferramentas Disponíveis, Criticidade do Software, Orçamento e Cronograma. Para complementar foram adicionados: Tipo de Projeto e Plataforma Tecnológica [11].

As entradas foram organizadas em categorias da seguinte forma:

- *Tamanho da Equipe:* PP – Muito Pequena (de 1 a 6 pessoas), P – Pequena (de 7 a 20 pessoas), M – Média (de 21 a 50 pessoas), G – Grande (de 51 a 100 pessoas) e GG – Muito Grande (mais de 100 pessoas);

- *Distribuição geográfica da equipe:* MS – Mesma sala, MP – Mesmo prédio e salas diferentes, MCO – Mesma cidade, prédio diferente e mesma organização, MCOOD – Mesma cidade e organização diferente, CD – Cidades diferentes;
- *Tamanho do projeto:* PP – Muito Pequena (até R\$ 50.000), P – Pequena (de R\$ 50.000 a R\$ 150.000), M – Média (de R\$ 150.000 a R\$ 1.000.000), G – Grande (de R\$ 1.000.000 a R\$ 3.000.000) e GG – Muito Grande (mais de R\$ 3.000.000);
- *Tipo de projeto:* Jogo, Financeiro, Automação Comercial, ERP, Logística, Científico, Governamental e Outros;
- *Tecnologia utilizada:* J2EE, Asp, Php, Java, Delphi, .NET, C/C++ e Outros;
- *Experiência em processo, Experiência no domínio da aplicação e Experiência técnica:* SE – Sem experiência, 1P – Um projeto de experiência, 2a3P – De dois a três projetos de experiência, 4a5P – De quatro a cinco projetos de experiência e 5mP – Mais de 5 projetos.

4.1.2 Resultados dos Projetos (Saída da Rede)

Os riscos envolvidos em um projeto de software são muitos. No intuito de sistematizar estes riscos algumas taxonomias de riscos foram propostas. As dezenas de riscos levantados nestas taxonomias compõem um conjunto muito grande de informações que necessitariam de uma análise bastante detalhada dos projetos. Porém, o objetivo principal deste trabalho é apenas verificar a viabilidade da aplicação da MLP para avaliação de riscos. Assim foi feita a opção por analisar os efeitos dos riscos nos aspectos mais importantes de um projeto, conforme mostrado na Figura 4.1.

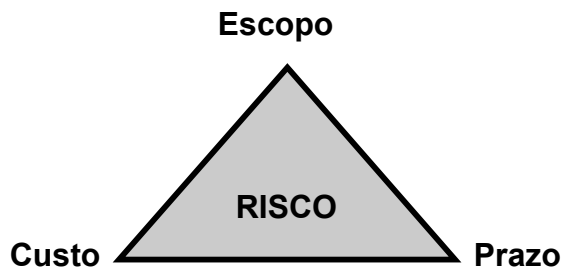


Figura 4.1. Pilares do gerenciamento de risco

Com o foco nos pilares do gerenciamento de riscos a caracterização dos resultados do projeto observou os seguintes aspectos: Finalização, Sobretrabalho, Retrabalho e a relação entre planejado/realizado em custo, prazo e escopo.

As saídas foram organizadas em categorias da seguinte forma:

- *Finalização, Sobretrabalho e Retrabalho:* Sim ou Não;
- *Prazo:* 10pct - até 10% de atraso, 25pct - de 10% a 25% de atraso, 50pct - de 25% a 50% de atraso, 100pct - de 50% a 100% de atraso e m100pct – mais de 100% de atraso;
- *Custo:* 10pct - até 10% de aumento no custo, 25pct - de 10% a 25% de aumento no custo, 50pct - de 25% a 50% de aumento no custo, 100pct - de 50% a 100% de aumento no custo e m100pct – mais de 100% de aumento no custo;
- *Escopo:* 10pct - até 10% de aumento no escopo, 25pct - de 10% a 25% de aumento no escopo, 50pct - de 25% a 50% de aumento no escopo, 100pct - de 50% a 100% de aumento no escopo e m100pct – mais de 100% de aumento no escopo;

4.1.3 Coleta de Dados

Com a finalidade de obter dados da forma mais objetiva possível foi disponibilizado um questionário na web e divulgado junto nos grupos da área de gerenciamento de projetos com o intuito de atingir os gerentes de projeto.

O questionário com catorze questões de múltipla escolha foi dividido em dois conjuntos de perguntas. As oito primeiras foram referentes à caracterização do

projeto e as seis seguintes sobre os resultados do projeto. Cada questionário respondido correspondia a um projeto específico.

O questionário foi disponibilizado durante o mês de maio e obteve respostas de 38 projetos, porém duas respostas foram descartadas por não apresentar as informações referentes à finalização do projeto em questão. Com isto, apenas 36 respostas foram aproveitadas.

No apêndice A encontram-se o questionário utilizado e o e-mail de divulgação da coleta de dados no apêndice B.

4.2 Prototipação da MLP

Para construção do protótipo da MLP foi utilizado o software Weka¹. Esta ferramenta foi desenvolvida pela Universidade de Waikato da Nova Zelândia. O Weka é um poderoso software que oferece ferramentas para o pré-processamento, classificação, regressão, clusterização, regras de associação e visualização. Desta forma se adequar à necessidade deste trabalho que buscou a criação de protótipos de redes na busca de uma topologia que apresente resultados satisfatórios.

No processo completo de aprendizagem os dados serão divididos em três conjuntos (treinamento, validação e teste). Esta divisão se faz necessário, pois existem necessidades além do treinamento. O conjunto de validação é necessário para evitar o overfitting, pois do contrario a rede pode “decorar” os dados e perder sua capacidade de generalização. E após o treinamento realizado (e validado) é necessário testar a rede para medir seu desempenho. Neste trabalho a distribuição entre os conjuntos será de 50% para treino, 25% para validação e 25% para testes.

As seções abaixo apresentam as etapas realizadas durante o processo de prototipação da MLP.

¹ Weka na WEB: <http://www.cs.waikato.ac.nz/ml/weka/>

4.2.1 Tratamento e pré-processamento dos Dados

Para alimentar a MLP os dados obtidos a partir do questionário foram transformados em vetores de dados com base na codificação apresentada nas seções 4.1.1 e 4.1.2. Estes dados foram tratados para assumir o formato de dados aceito pelo weka (arquivos .arff). Os quatro arquivos .arff produzidos se encontram nos apêndices.

Ao realizar uma análise preliminar a partir dos 36 vetores de dados obtidos temos os mapas de distribuição das variáveis envolvidas conforme figura 4.2. Esta figura mostra a proporção das saídas em cada parâmetro de entrada, o azul representa os projetos que foram finalizados, o vermelho são os projetos não finalizados.

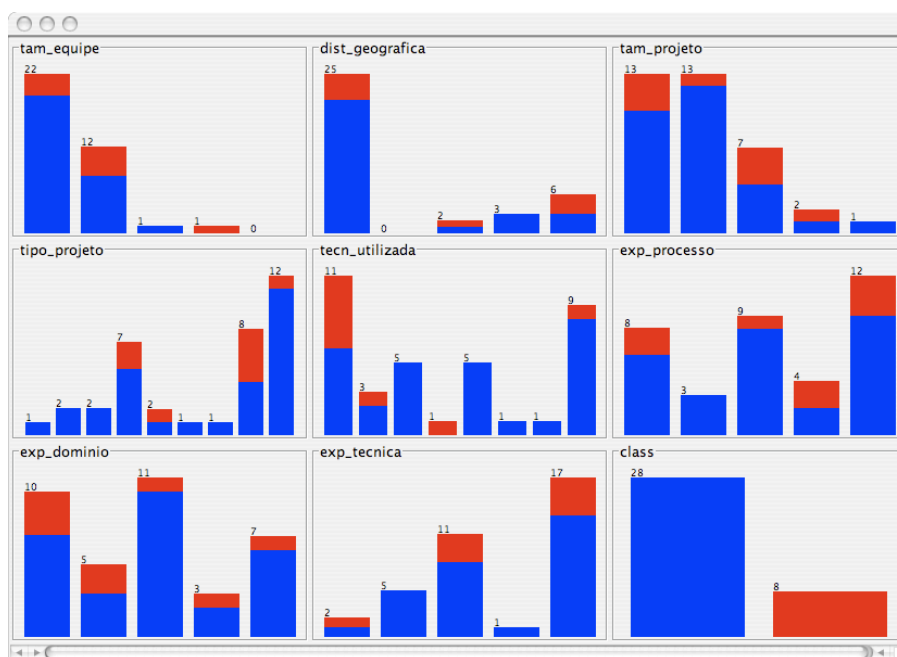


Figura 4.2. Distribuição dos dados de entrada coletados

Com uma análise nestes dados ficou evidenciado que as entradas referentes ao tamanho da equipe e tamanho do projeto estão concentradas em um subconjunto das entradas. Para compreender melhor o problema temos o seguinte detalhamento da situação:

- Tamanho da equipe: mais de 94% em apenas duas categorias (Muito pequena e Pequena);

- Tamanho do projeto: mais de 91% em três categorias (Muito pequeno, Pequeno e Médio).

Nesta situação de falta de dados para caracterizar projetos enquadrados nas categorias grandes e muito grandes os projetos com estas características foram excluídos e o estudo se concentrou nos projetos caracterizados como: muito pequenos, pequenos e médios. Com a restrição necessária 4 projetos foram excluídos restando 32 vetores de dados. Assim foram divididos em 16 para treinamento e 8 para validação.

4.2.2 Topologia e parâmetro de aprendizagem da MLP

As redes MLP foram construídas com três camadas e a quantidade de neurônio em cada camada é apresentada na tabela 4.6.

Tabela 4.1. Quantidade de Neurônio por camada em cada MLP

Descrição da MLP	Camada de Entrada	Camada Oculta	Camada de Saída
Finalização	41	23	2
Prazo	41	29	6
Custo	41	29	6
Escopo	41	29	6

A rede foi submetida a treinamento supervisionado com a validação cruzada como critério de parada e o limite superior de 1000 épocas. A taxa de aprendizagem utilizada foi de 0,1 e o momento de 0,2.

4.2.3 Resultados da MLP

Para análise dos resultados obtidos quatro tabelas foram organizadas com o resultado de cada vetor do conjunto de teste em cada MLP. Esta tabela é composta por cinco colunas, sendo elas: uma coluna com o valor constante do conjunto de dados (valor real), uma coluna com o valor que a MLP indicou para aquele projeto (valor proposto), uma coluna indicando se o valor proposto está correto (Situação) e uma coluna onde o percentual calculado pela MLP para cada possibilidade de saída é apresentada (Distribuição de Probabilidade).

Tabela 4.2. Apresentação dos resultados da MLP para finalização

Valor Real	Valor MLP	Situação	Distribuição de Probabilidade	
			S	N
S	S	Certo	99,9%	0,1%
N	S	Errado	99,9%	0,1%
S	S	Certo	96,7%	3,3%
S	S	Certo	99,9%	0,1%
N	S	Errado	100%	0%
N	N	Certo	10,5%	89,5%
S	S	Certo	99,9%	0,1%
N	S	Errado	100%	0%

A MLP construída para avaliação da finalização do projeto, ou seja, se o projeto será finalizado ou não, apresentou 62,5% de acerto. Como se trata de apenas duas possibilidades de saída (sim e não) o resultado foi ligeiramente superior aos 50% de uma escolha ao acaso. O conjunto de dados limitado impossibilita a conclusão sobre a eficiência da MLP para este fim.

Tabela 4.3. Apresentação dos resultados da MLP para prazo

Valor Real	Valor MLP	Situa.	Distribuição de Probabilidade					
			S	10pct	25pct	50pct	100pct	m100pct
10pct	100pct	Errado	4,1%	0,1%	0,4%	1,5%	93,4%	0,5%
S	S	Certo	72,3%	0%	21%	1,7%	4,4%	0,6%
S	25pct	Errado	5,3%	0%	82,8%	0,1%	11,2%	0,5%
m100pct	25pct	Errado	27,2%	0,4%	71,5%	0,5%	0%	0,4%
10pct	10pct	Certo	0%	57,5%	24,6%	17,7%	0%	0,2%
10pct	10pct	Certo	0%	51%	14,3%	34,3%	0%	0,4%
100pct	100pct	Certo	0,1%	0,1%	10%	2,5%	86,7%	0,7%
S	S	Certo	79,7%	13,9%	0,6%	0%	5,2%	0,5%

Na MLP para avaliação de prazo o resultado encontrado foi um acerto de 62,5% no atraso previsto para ao projeto. Mesmo com o resultado igual à rede para finalização a situação aqui é completamente diferente, pois são seis possibilidades de classificação.

Tabela 4.4. Apresentação dos resultados da MLP para custo

Valor Real	Valor MLP	Situa.	Distribuição de Probabilidade					
			S	10pct	25pct	50pct	100pct	m100pct
S	S	Certo	88,3%	8,6%	0%	1,9%	0,9%	0,3%
m100pct	S	Errado	84,9%	0,1%	0,1%	14%	0,7%	0,3%
S	S	Certo	98,6%	0,3%	0%	0,6%	0,2%	0,3%
S	50pct	Errado	0,9%	25,2%	7,3%	62,4%	2,4%	1,7%
S	10pct	Errado	5,1%	62,5%	0%	0,4%	31,5%	0,5%
S	S	Certo	57,5%	32,7%	0%	0,3%	9,3%	0,3%
50pct	50pct	Certo	4,9%	10,4%	2,5%	77,4%	3%	1,7%
S	S	Certo	93,4%	0%	0%	5,5%	0,7%	0,3%

A avaliação de custo apresentou os mesmos resultados da MLP para prazo, ou seja, 62,5% das predições efetuadas para a evolução do custo do projeto foram acertadas.

Tabela 4.5. Apresentação dos resultados da MLP para escopo

Valor Real	Valor MLP	Situa.	Distribuição de Probabilidade					
			S	10pct	25pct	50pct	100pct	m100pct
S	S	Certo	86,5%	0,1%	5,4%	4,7%	2,9%	0,4%
25pct	25pct	Certo	1,9%	0,1%	92,1%	0,3%	5,1%	0,5%
25pct	25pct	Certo	7,7%	0,1%	88,8%	0,2%	2,9%	0,4%
25pct	S	Errado	60,8%	0%	0,7%	2,1%	35,7%	0,6%
25pct	25pct	Certo	0,1%	2,9%	93,5%	1,1%	1,9%	0,5%
25pct	25pct	Certo	0,2%	0,7%	95,8%	0,8%	1,9%	0,5%
25pct	25pct	Certo	3,6%	0%	90,7%	0,2%	3,1%	0,5%
100pct	25pct	Errado	2,7%	6%	90%	0,8%	0,2%	0,4%

A MLP para avaliação do escopo apresentou o resultado de 75% de acerto em sua predição. Com este resultado esta rede obteve o melhor desempenho de todas.

4.3 Avaliação dos resultados MLP

A redes MLP construídas para prazo, custo e escopo apresentaram resultados satisfatórios uma vez que três redes apresentaram uma taxa de acerto de 62,5% e uma com 75%. A MLP para finalização mesmo com o resultado igual às MLP de prazo e custo seu resultado não possibilita nenhuma conclusão uma vez que esta rede possui apenas duas alternativas de saída. Desta forma o resultado obtido fica muito próximo de uma escolha ao acaso e como o conjunto de dados de

teste é pequeno não existe a garantia alguma que este resultado não seja fruto do acaso. Com um consumo de dados mais significativos do problema maiores inferências estatísticas poderam se realizadas.

Na análise dos resultados para prazo, custo e escopo podemos pegar o exemplo da rede para avaliação do prazo. Aqui existem 6 alternativas de resposta, ou seja, se os resultados fossem bem distribuídos a probabilidade de acertar sorteando um valor seria de menos de 17%. O resultado de 62,5% é muito superior a uma escolha ao acaso.

Outra possibilidade seria escolher a saída que mais aparece, no caso do prazo a probabilidade de acertar seria de um pouco mais de 31%. Mesmo nestas condições o desempenho da rede é ainda muito melhor.

Com base nesta avaliação o resultado encontrado demonstra indícios muito fortes da existência de padrões de associação entre as características dos projetos e os resultados obtidos. Com isto a utilização de Redes Neurais Artificiais se mostra como uma boa alternativa para construção de ferramenta de suporte ao gerenciamento de riscos.

Um ponto importante de avaliação é a quantidade limitada dos dados, este estudo encontrou dificuldades na coleta de dados, devido à resistência dos profissionais de TI em fornecer as informações. Os motivos das resistências são dos mais variados, mas passam principalmente pela falta de registro das informações dos projetos passados e alegação da falta de tempo para responder a pesquisa. O conjunto limitado dos dados limitou a exploração do potencial das redes MLP. Com a continuação deste estudo e conseqüentemente à ampliação de dados, resultados melhores certamente acompanharão o estudo.

4.3.1 Fatores positivos e negativos das técnicas

É perceptível pelo apresentado na seção 3.1.1 que a vantagem da utilização de árvores de decisão se dá pela sua simplicidade sendo uma boa ferramenta para resolver problemas de baixa complexidade e com baixa utilização de recursos

computacionais. Porém a avaliação de riscos não é um problema simples, desta forma a aplicação de outras técnicas tende a obter resultados melhores.

A técnica de RBC é uma técnica com um bom potencial para resolver os problemas pertinentes ao gerenciamento de riscos. Além de indicar os casos que já ocorreram servindo de fonte de identificação de riscos em potencial ainda apresenta possíveis soluções. Uma limitação desta técnica reside na etapa de assessoramento da situação, pois é fundamental que seja feita uma boa definição de descritores relevantes ao problema. Como os problemas podem ter um alto grau de complexidade, a identificação destes descritores pode não ser uma tarefa trivial, criando a possibilidade do sistema não possuir uma boa representação do problema, o que inevitavelmente acarretará em falha na identificação.

A Lógica Fuzzy é uma técnica consagrada no suporte à tomada de decisão. Embora nenhuma ferramenta específica para avaliação de riscos com suporte da Lógica Fuzzy tenha sido encontrada as demais aplicações desta técnica apontam o seu potencial no suporte a tomada de decisão. Uma limitação na aplicação de Lógica Fuzzy na avaliação de riscos é a necessidade de um conhecimento significativo deste assunto por parte dos usuários da ferramenta. Esta necessidade é um limitador para sua aplicação em larga escala. Pois a aplicação de uma ferramenta baseada em Fuzzy terá que vir acompanhada da capacitação de gestores. Isto é um limitador para incorporação por pequenas e médias empresas. Outra dificuldade consiste na alta subjetividade envolvida no processo de definição do grau de pertinência algo que também limita a aplicação em larga escala.

As Redes Bayesianas são ferramentas com uma excelente resultados em diversas áreas, inclusive com ferramentas comerciais desenvolvidas para a avaliação de riscos. A dificuldade da utilização das Redes Bayesianas se dá pela necessidade de se conhecer as probabilidades condicionais. Considerando que as organizações em geral possuem um registro muito precário dos fatores relevantes de um projeto. Então, conseguir levantar o nível de interação entre estes fatores nas pequenas e médias organizações é pouco provável.

As redes MLP são ferramentas muito poderosas que possibilitam resolver problemas com alto grau de complexidade. No domínio do problema do

gerenciamento de risco o MLP se mostrou com um bom potencial para avaliação de riscos. A utilização de MLP fornece a ferramenta à capacidade de absorver parte da análise do processo. Isto possibilita que o usuário abstraia os mecanismos internos de funcionamento da ferramenta e se concentre em outros aspectos do projeto. Uma grande limitação desta técnica é a necessidade de uma quantidade considerável de dados sobre o histórico dos projetos, que nem sempre estão disponíveis.

4.4 Resumo do Capítulo

Este capítulo tem a finalidade de descrever todas atividades desenvolvidas para implementar o protótipo de rede MLP proposta neste trabalho. Além da prototipação as atividades de Caracterização do projeto, da Caracterização da Finalização e da Coleta de Dados são descritas neste capítulo. A conclusão deste capítulo é feita com a análise dos resultados obtidos e a comparação da utilização da MLP para avaliação de riscos em relação às demais técnicas de computação inteligentes apresentadas neste trabalho.

Capítulo 5

Conclusão e Trabalhos Futuros

Este trabalho visou à avaliação da viabilidade e eficiência do uso das redes MLP na avaliação de riscos em projeto de software. Com base nos resultados deste trabalho revela-se que existe um grande potencial a ser explorado. Estes resultados são mais animadores dado o contexto atual da Gerência de Projetos.

Este trabalho deve ser visto como um estudo preliminar que a sua continuidade poderá resultar na construção de uma ferramenta robusta para suporte ao gerenciamento de riscos.

5.1 Contribuição

A principal contribuição deste trabalho é a indicação da viabilidade em utilizar as redes MLP para avaliação de riscos em projetos de software. O resultado de no mínimo 62,5% encontrado aponta o potencial da técnica.

Este trabalho demonstra que a continuidade deste estudo possibilitará o desenvolvimento de um módulo para o OpenmPRIME. A inclusão de uma ferramenta com base em MLP possibilitará um melhor planejamento e acompanhamento dos riscos. Com a elaboração de estudos e ferramentas como a aqui proposta é de se esperar uma melhora no cenário do desenvolvimento de software.

A criação de técnicas que possibilitem a utilização dos dados históricos dos projetos por si só apresenta o potencial de estimular a sistematização das informações relevantes no andamento dos projetos. Além disto o registro das informações dos projetos possibilita o amadurecimento das organizações por meio de sua própria experiência. O valor que estas informações pode ser ampliado se disponibilizadas para estudos externos que trazem benefícios para todos.

5.2 Trabalhos Relacionados

Não foi encontrada na literatura ferramenta utilizando redes MLP para avaliação de risco. Porém um trabalho se relaciona a este estudo por utilizar uma técnica de computação inteligente na identificação de riscos. O trabalho em questão foi apresentado pelo aluno Arthur Lins como Projeto Final do Curso de Engenharia da Computação na Universidade de Pernambuco sobre o título: CBR Risk Method [11].

5.3 Dificuldades Encontradas

Neste trabalho duas questões restringiram a possibilidade de um estudo mais conclusivo a cerca da aplicabilidade do uso de Redes Neurais Artificiais na avaliação de risco. A principal limitação deste trabalho é decorrente da dificuldade na coleta de dados junto aos profissionais de TI. Mesmo sendo feito um grande esforço para conseguir os dados que subsidiaram este estudo, a quantidade atingida dificultou o treinamento da rede. A situação se agravou com relação aos dados coletados referentes a grandes projetos. Isto resultou na restrição do universo da pesquisa uma vez que não havia dados suficientes para estudar o comportamento destes projetos. Outra dificuldade relevante vem da escassez de resultados práticos sobre o desempenho das técnicas existentes para avaliação de riscos. Durante a pesquisa foram encontradas propostas de técnicas e ferramentas para este fim, porém nenhum estudo consistente comprovando o desempenho destas foi encontrado. Isto provocou uma grande dificuldade em comparar as técnicas existentes com a proposta da utilização de Redes Neurais Artificiais.

As dificuldades encontradas neste estudo não trouxeram apenas impactos negativos, mas possibilitaram compreender aspectos importantes a cerca da gerência de projetos. Uma lição importante foi a valorização do registro dos históricos dos projetos. A necessidade de uma mudança de comportamento por parte dos gestores dos projetos para possibilitar a aprendizagem com suas próprias experiências foi outra lição valiosa. Além disto o mais importante foi o entendimento

de que área de gerência de riscos carece de pesquisas que contribuam para uma mudança na realidade do desenvolvimento de software.

5.4 Trabalhos Futuros

Este trabalho apontou a necessidade de realizar algumas atividades visando as melhorias necessárias para a implementação profissional da técnica. Desta forma as seguintes ações são sugeridas:

- Adição de informações de caracterização – Deve se estudar o levantamento de novas características dos projetos para uma avaliação mais eficiente.
- Estudos comparativo com outras RNAs – Realizar estudo comparativo do desempenho com outras RNAs, tais como: Wavelets, RBF ou mesmo outras topologias de MLP.
- Ampliação das respostas – Estudo para ampliar as saídas da rede a fim de avaliar todos os riscos levantados pelas taxonomias de riscos mais relevantes.
- Ampliação para suporte a múltiplos projetos – Inclusão de aspectos relativos à interação entre os projetos em execução na organização com a finalidade de enriquecer a avaliação. Assim será possível avaliar os riscos de forma mais contextualizada.
- Implementação do módulo de Avaliação de Riscos para o OpenmPrime – Desenvolvimento de um módulo específico para avaliação de risco no sistema.

Bibliografia

- [1] Agile Manifesto, <http://agilemanifesto.org/>, acessado em 12 de Setembro de 2008.
- [2] BOEHM, B. W. **Software Risk Management: Principles and Practices**. IEEE Software, vol. 8, n. 1, p. 32-40, 1993.
- [3] BRAGA, A. P.; CARVALHO, A. C. P. L. F. e LUDEMIR, T. B. **Redes Neurais Artificiais**, vol. 1, cap. 6, p. 30, 2003.
- [4] BRUNETTO, C., et al. **Uma comparação entre Redes Neurais Wavelet, LMS, MLP e RBF para classificação de DPOC**. Congresso de Matemática e suas Aplicações, Foz do Iguaçu, Brasil, 2006.
- [5] CARR, M. J., et al. **Taxonomy-Based Risk Identification**. Relatório Técnico. CMU/SEI-93-TR-6. Software Engineering Institute, Carnegie Mellon University, Pittsburgh, USA, 1993.
- [6] COELHO, C. C. **MAPS: um Modelo de Adaptação de Processos de Software**. Dissertação de Mestrado, Universidade Federal de Pernambuco, Recife, 2003.
- [7] FUJIMOTO, R. Y. **Diagnóstico Automático de Defeitos em Rolamentos Baseados em Lógica Fuzzy**. Dissertação de Mestrado, Universidade de São Paulo, São Paulo, Brasil, 2005.
- [8] GADI, M. F. A. **Uma comparação de métodos de classificação aplicados à detecção de fraude em cartões de crédito**. Dissertação de Mestrado, Universidade de São Paulo, São Paulo, 2008.
- [9] GUSMÃO, C. M. G. **Um Modelo de Processo de Gestão de Riscos para Ambientes de Múltiplos. Projetos de Desenvolvimento de Software**. Tese de Doutorado, Universidade Federal de Pernambuco, Recife, 2007.

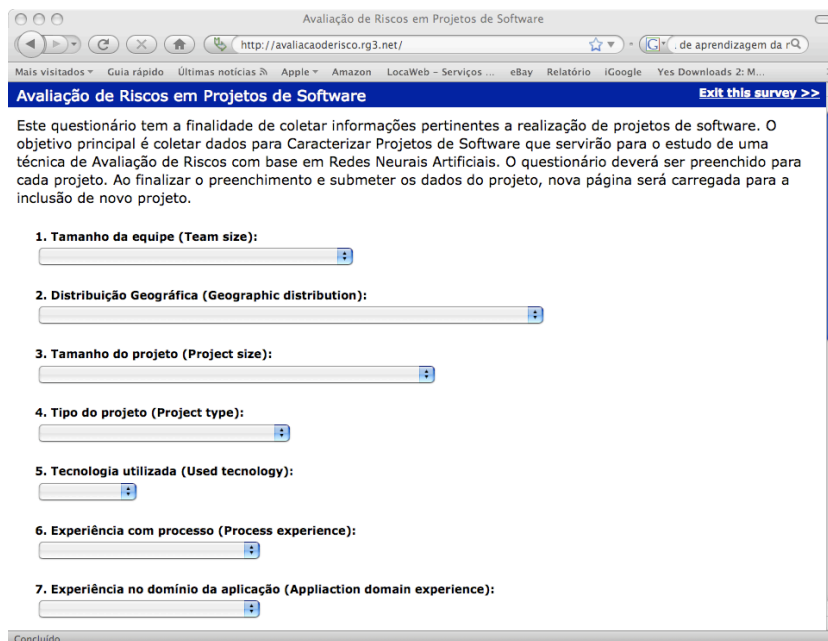
- [10] LEOPOLDINO, C. B. **Avaliação de Riscos em Desenvolvimento de Software**. Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.
- [11] LINS, A. **CBR Risk Method**. Trabalho de Graduação. Universidade de Pernambuco, Recife, Brasil, 2007.
- [12] LOESCH, C. e SARI, S. T. **Redes neurais artificiais: fundamentos e modelos**. FURB, Blumenau, 1996.
- [13] LUU, V. T., et al. **Quantifying schedule risk in construction projects using Bayesian belief networks**. International Journal of Project Management, n. 27, p. 39-50, 2009.
- [14] MACHADO, C. A. F. **A-RISK: Um método para identificar e quantificar riscos de prazo Em projeto de desenvolvimento de software**. Dissertação de Mestrado, Pontifícia Universidade Católica do Paraná, Curitiba, Brasil, 2002.
- [15] MAXIMIANO, A. C. M. **Administração de Projetos: Como transformar idéias em resultados**. São Paulo, Atlas, 1997. 196 p.
- [16] MILARÉ, C. R. **Extração de conhecimento de Redes Neurais Artificiais utilizando sistemas de aprendizado simbólico e algoritmos genéticos**. Tese de Doutorado, Universidade de São Paulo, São Carlos, 2003.
- [17] PROJECT MANAGEMENT INSTITUTE. **Um Guia do Conjunto de Conhecimento em Gerenciamento de Projetos (Guia PMBOK®)**. Terceira Edição, 2004.
- [18] QUEIROZ, C. N. **Redes Bayesianas no Gerenciamento Mensuração de Riscos Operacionais**. Dissertação de Mestrado, Universidade de São Paulo, São Paulo, 2008.
- [19] SILVEIRA, F. P. e KNOB, F. F. **RiskFree: Uma ferramenta de apoio à gerência de riscos em projetos de software**. Trabalho de Conclusão de

Curso, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brasil, 2005.

- [20] SOFTWARE ENGINEERING INSTITUTE. **CMMI Capability Maturity Model Integration**. v. 1.1 Pittsburgh, PA. Software Engineering Institute, Carnegie Mellon University. USA, 2001.
- [21] SOUZA, C. A. **Teoria de conjuntos fuzzy e regressão logística na tomada de decisão para realização de cintilografia das paratireóides**. Dissertação de Mestrado, Universidade de São Paulo, São Paulo, Brasil, 2007.
- [22] STANDISH GROUP. **CHAOS report**. 586 Old Kings Highway, Dennis, MA 02638, USA, 2004.
- [23] TRIGO, T. **Avaliação Experimental do CBR Risk Method**. Trabalho de Graduação, Universidade de Pernambuco, Recife, Brasil, 2007.
- [24] TURING, A. M (1950). **Computing machinery and intelligence**. New Series, vol. 59, n. 236, p. 433-460.
- [25] WITTEN, I. H e FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**. Morgan Kaufmann Press, 2005.

Apêndice A

Questionário de Coleta de Dados



Avaliação de Riscos em Projetos de Software

Este questionário tem a finalidade de coletar informações pertinentes a realização de projetos de software. O objetivo principal é coletar dados para Caracterizar Projetos de Software que servirão para o estudo de uma técnica de Avaliação de Riscos com base em Redes Neurais Artificiais. O questionário deverá ser preenchido para cada projeto. Ao finalizar o preenchimento e submeter os dados do projeto, nova página será carregada para a inclusão de novo projeto.

1. Tamanho da equipe (Team size):

2. Distribuição Geográfica (Geographic distribution):

3. Tamanho do projeto (Project size):

4. Tipo do projeto (Project type):

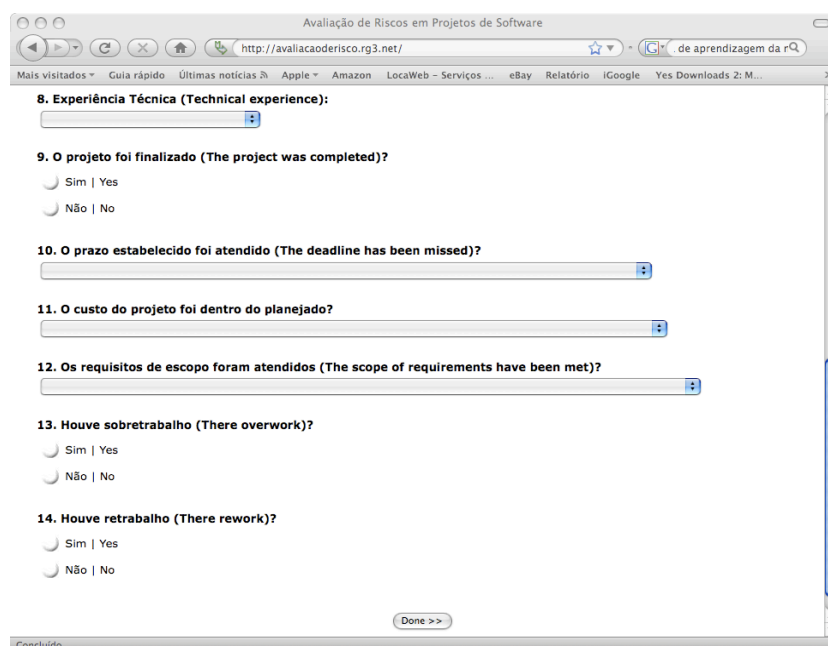
5. Tecnologia utilizada (Used tecnology):

6. Experiência com processo (Process experience):

7. Experiência no domínio da aplicação (Appliaction domain experience):

Concluído

Figura A.1. Parte superior do questionário de coleta de dados



Avaliação de Riscos em Projetos de Software

8. Experiência Técnica (Technical experience):

9. O projeto foi finalizado (The project was completed)?
 Sim | Yes
 Não | No

10. O prazo estabelecido foi atendido (The deadline has been missed)?

11. O custo do projeto foi dentro do planejado?

12. Os requisitos de escopo foram atendidos (The scope of requirements have been met)?

13. Houve sobretrabalho (There overwork)?
 Sim | Yes
 Não | No

14. Houve retrabalho (There rework)?
 Sim | Yes
 Não | No

Done >>

Concluído

Figura A.2. Parte inferior do questionário de coleta de dados

Apêndice B

E-mail de divulgação

Coleta de dados para Pesquisa na área de Identificação de Riscos

Mesmo com a evolução dos computadores, das técnicas e ferramentas nos últimos anos, a produção de software confiável, correto e entregue dentro dos prazos e custos estipulados ainda é muito difícil.

Os relatórios apresentados pelo Standish Group (Chaos Report) nos últimos 15 anos trazem resultados nada animadores. No estudo com mais de 175.000 projetos o primeiro resultado foi publicado em 1994. Neste ano apenas 16% dos projetos estudados foram finalizados com sucesso. Dez anos após, em 2004 o relatório apontou uma melhoria mais o resultado ainda é muito inexpressivo. Apenas 25% dos projetos foram finalizados com sucesso.

Diante deste cenário onde 75% dos projetos apresentam algum problema grave que impacte no custo, prazo ou escopo do projeto. Assim o gerenciamento de riscos nitidamente aparece como uma ferramenta vital para o gerente do projeto.

Motivado pela necessidade da comunidade de desenvolvimento de software o grupo de pesquisa PROMISE atualmente estuda o desenvolvimento de uma ferramenta aberta para gerenciamento de riscos (OpenmPrime).

Dentro desta perspectiva foi iniciado o desenvolvimento de técnica para identificação de riscos a partir de Redes Neurais Artificiais. Os dados para treinamento e avaliação desta proposta serão coletados através do preenchimento de questionário (<http://avaliacaoderisco.rg3.net/>). O preenchimento é rápido e objetivo.

Agradecemos antecipadamente a colaboração.

Dúvidas ou sugestões entrar em contato com
Raphael D´Castro (raphaeldcastro@gmail.com)