Pontifícia Universidade Católica
do Rio de Janeiro

Bernardo Lins de Albuquerque Compagnoni

Development of Prediction Models for the Quality of Spoken Dialogue Systems

Dissertação de Mestrado

Dissertation presented to the Postgraduate Program in Electrical Engineering of the Departamento de Engenharia Elétrica, PUC-Rio as partial fulfillment of the requirements for the degree of Mestre em Engenharia Elétrica.

Advisor: Prof. Cristiano Augusto Coelho Fernandes

Rio de Janeiro

April 2011

Pontifícia Universidade Católica
do Rio de Janeiro

Bernardo Lins de Albuquerque Compagnoni

Development of Prediction Models for the Quality of Spoken Dialogue Systems

Dissertation to the presented Postgraduate Program in Electrical Engeneering, of the Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio, as partial fulfillment of the requirements for the degree of Mestre

Prof. Cristiano Augusto Coelho Fernandes
Advisor
Departamento de Engenharia Elétrica – PUC-Rio

Prof. Fernando Gil Vianna Rsende
UFRJ

Prof. Abraham Alcaim
Centro de Estudos em Telecomunicações – PUC-Rio

Prof. José Eugenio Leal
Coordinator of the Centro Técnico Científico–
PUC-RIO

Rio de Janeiro, 04 de abril de 2011

Bernardo Lins de Albuquerque Compagnoni

Graduated in Electrical Engineering from PUC-Rio in 2008. Participated in a double degree program with the University of Braunschweig (Germany), between 2004 and 2006. Worked at Deutsche Telekom Laboratories in Berlin in 2006, where he did research on the topic of this dissertation. After graduating, he joined the Post-Graduate Program in Electrical Engineering from PUC-Rio to obtain a Masters degree

Bibliographic data

Compagnoni, Bernardo Lins de Albuquerque

Development of prediction models for the quality of spoken dialogue systems / Bernardo Lins de Albuquerque Compagnoni; advisor: Cristiano Augusto Coelho Fernandes. – 2011.

137 f. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2011.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Reconhecimento de voz. 3. Linguagem falada. 4. Spoken dialogues systems. 5. Regressão linear. 6. Árvores de regressão. 7. Árvores de classificação. 8. Redes neurais. 9. Avaliação de performance de sistemas. I. Fernandes, Cristiano Augusto Coelho. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

À minha filha Alice

Agradecimentos

À minha família. Ao meu pai Luiz, minha mãe Maria Regina, meu irmão Tiago e minha filha Alice, além de todos os meus amigos, pelo carinho, amor e apoio neste desafio. Sem eles seria impossível dar este passo.

Ao orientador Cristiano Fernandes, pela oportunidade concedida e confiança nas responsabilidades envolvidas. Agradeço pela excelente orientação em todas as etapas deste trabalho.

Aos orientadores Sebastian Möller e Tim Fingscheid, pela motivação e oportunidades concedidas durante a minha estadia na Alemanha.

À PUC-Rio, pelos auxílios concedidos e pelo ótimo ambiente de estudo.

Resumo

Compagnoni, Bernardo Lins de Albuquerque; Fernandes, Cristiano Augusto Coelho (Orientador). **Desenvolvimento de Modelos para Previsão de Qualidade de Sistemas de Reconhecimento de Voz**. Rio de Janeiro, 2011. 137p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Spoken Dialogue Systems (SDS's) são sistemas baseados em computadores desenvolvidos para fornecerem informações e realizar tarefas utilizando o diálogo como forma de interação. Eles são capazes de reconhecimento de voz, interpretação, gerenciamento de diálogo e são capazes de ter uma "voz" como saída de dados, tentando reproduzir uma interação natural falada entre um usuário humano e um sistema. SDS's provém diferentes serviços, todos através de linguagem falada com um sistema. Mesmo com todo o desenvolvimento nesta área, há escassez de informações sobre como avaliar a qualidade de tais sistemas com o propósito de otimização do mesmo. Com dois destes sistemas, BoRIS e INSPIRE, usados para reservas de restaurantes e gerenciamento de "casas inteligentes", diversos experimentos foram conduzidos no passado, onde tais sistemas foram utilizados para resolver tarefas específicas. Os participantes avaliaram a qualidade do sistema em uma série de questões. Além disso, todas as interações foram gravadas e anotadas por um especialista.O desenvolvimento de métodos para avaliação de performance é um tópico aberto de pesquisa na área de SDS's. Seguindo a idéia do modelo PARADISE (PARAdigm for DIalogue System Evaluation – desenvolvido pro Walker e colaboradores na AT&T em 1998), diversos experimentos foram conduzidos para desenvolver modelos de previsão de performance de sistemas de reconhecimento de voz e linguagem falada. O objetivo desta dissertação de mestrado é desenvolver modelos que permitam a previsão de dimensões de qualidade percebidas por um usuário humano, baseado em parâmetros instrumentalmente mensuráveis utilizando dados coletados nos experimentos realizados com os sistemas BoRIS e INSPIRE , dois sistemas de reconhecimento de voz (o primeiro para busca de restaurantes e o segundo para Smart Homes). Diferentes algoritmos serão utilizados para análise (Regressão linear, Árvores de Regressão, Árvores de Classificação e Redes Neurais) e para cada um dos algoritmos, uma ferramenta diferente será programada em MATLAB, para poder servir de base para análise de experimentos futuros, sendo facilmente modificado para sistemas e parâmetros novos em estudos subsequentes.A idéia principal é desenvolver ferramentas que possam ajudar na otimização de um SDS sem o envolvimento direto de um usuário humano ou servir de ferramenta para estudos futuros na área.

Palavras-chave

Engenharia Elétrica; Reconhecimento de Voz; Linguagem Falada; Spoken Dialogue Systems; Regressão linear; Árvores de Regressão; Árvores de Classificação; Redes Neurais; Avaliação de Performance de Sistemas.

Abstract

Compagnoni, Bernardo Lins de Albuquerque; Fernandes, Cristiano Augusto Coelho (Advisor). **Development of Prediction Models for the Quality of Spoken Dialogue Systems**. Rio de Janeiro, 2011. 137p. MSc Dissertation – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Spoken Dialogue Systems (SDS´s) are computer-based systems developed to provide information and carry out tasks using speech as the interaction mode. They are capable of speech recognition, interpretation, management of dialogue and have speech output capabilities, trying to reproduce a more or less natural spoken interaction between a human user and the system. SDS´s provide several different services, all through spoken language. Even with all this development, there is scarcity of information on ways to assess and evaluate the quality of such systems with the purpose of optimization. With two of these SDS´s ,BoRIS and INSPIRE, (used for Restaurant Booking Services and Smart Home Systems), extensive experiments were conducted in the past, where the systems were used to resolve specific tasks. The evaluators rated the quality of the system on a multitude of scales. In addition to that, the interactions were recorded and annotated by an expert. The development of methods for performance evaluation is an open research issue in this area of SDS´s. Following the idea of the PARADISE model (PARAdigm for DIalogue System Evaluation model, the most well-known model for this purpose (developed by Walker and co-workers at AT&T in 1998), several experiments were conducted to develop predictive models of spoken dialogue performance. The objective of this dissertation is to develop and assess models which allow the prediction of quality dimensions as perceived by the human user, based on instrumentally measurable variables using all the collected data from the BoRIS and INSPIRE systems. Different types of algorithms will be compared to their prediction performance and to how generic they are. Four different approaches will be used for these analyses: Linear regression, Regression Trees, Classification Trees and Neural Networks. For each of these methods, a different tool will be programmed using MATLAB, that can carry out all experiments from this work and be easily modified for new experiments with data from new systems or new variables on future studies. All the used MATLAB programs will be made available on the attached CD with an "operation manual" for future users as well as a guide to modify the existing programs to work on new data. The main idea is to develop tools that would help on the optimization of a spoken dialogue system without a direct involvement of the human user or serve as tools for future studies in this area.

Keywords

Electrical Engineering; Speech Recognition; Spoken Language; Spoken Dialogue Systems; Linear Regression; Regression Trees; Classification Trees; Neural Networks; System Performance Evaluation.

Summary

List of diagrams and figures

List of Tables