

APLICAÇÃO DE SELEÇÃO DE CARACTERÍSTICAS PARA MELHORIA DE UM SISTEMA SEMIAUTOMÁTICO DE AVALIAÇÃO DE QUESTÕES DISCURSIVAS

Marcos A. Spalenza

Orientador: Elias de Oliveira

Co-Orientadora: Márcia G. de Oliveira



UFES

PPGI - Programa de Pós Graduação em Informática - Mestrado em Informática
Universidade Federal do Espírito Santo

Agenda

- Introdução
 - Motivação
 - Objetivos
- Avaliação Assistida por Computação
- Avaliação de Respostas Discursivas
- Experimentos e Resultados
 - Métricas
 - Algoritmos
 - Bases de Dados
- Conclusões e Trabalhos Futuros

Introdução

As verificações de aprendizado são fundamentais para o ensino para caracterização do desempenho, aperfeiçoamento pedagógico e reformulação do processo de ensino-aprendizagem Barreira et.al. (2006).

Segundo Oliveira e Santos (2005) o papel da avaliação é diagnosticar, apreciar e verificar a proficiência dos alunos para que o professor atue no processo de formação de modo a consolidar o aprendizado.

Como influente modelo de verificação pedagógica durante a formação do aprendiz, devemos destacar as atividades discursivas de leitura e interpretação textual como excelente exercício cognitivo para os alunos. Assim, nesse modelo de atividade reúnem-se diversos esforços para incentivo à escrita e leitura, como o PNLL¹.



¹ Plano Nacional do Livro e Leitura / Ministério da Cultura - Ministério da Educação - PNLL/MinC-MEC

Introdução - Motivação

Apesar da importância o processo de correção de atividades é de alto custo ao tempo desse profissional. Sendo o excesso de tarefas um grande causador de problemas de saúde dos professores Assunção e Oliveira (2009).

Enquanto isso, aos alunos, a falta de prática com as atividades causa defasagem nas habilidades básicas, tal como a leitura e escrita. À esse déficit, é atribuído graves problemas de aprendizagem e baixo rendimento acadêmico Martens (2016).

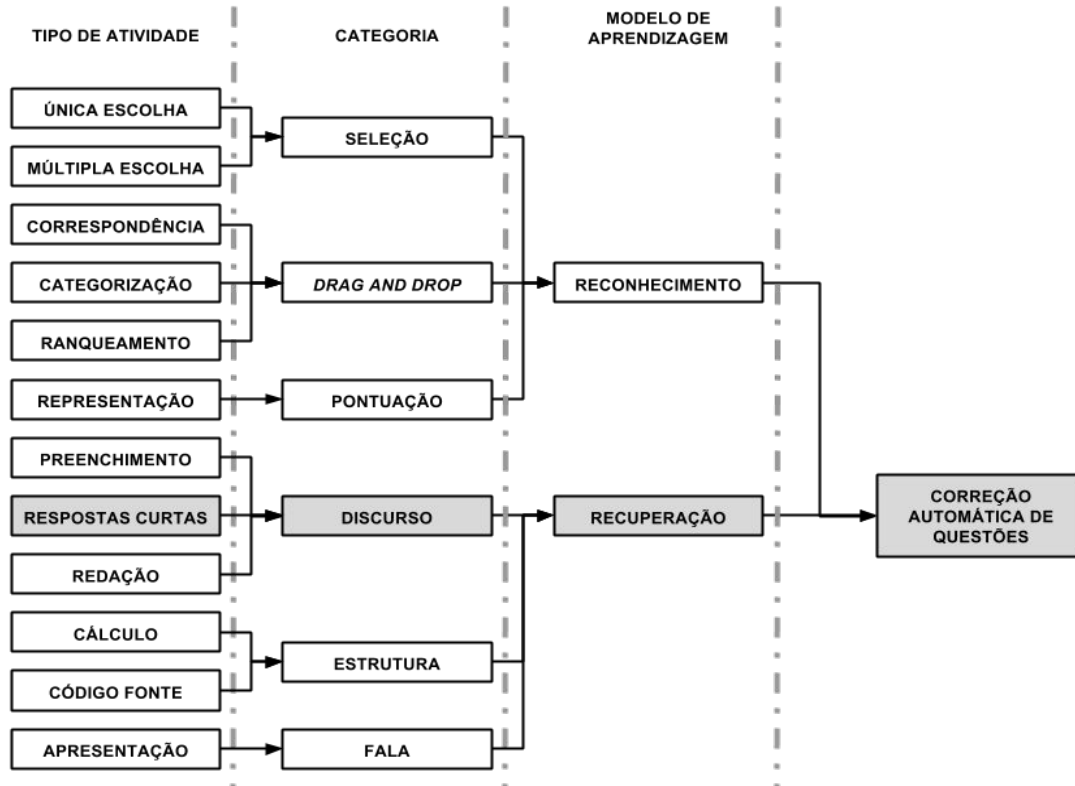
Conforme Lima e Giraffa (2006), a introdução de sistemas computacionais como recurso pedagógico torna o processo de ensino e de aprendizagem mais efetivo, incentivando a interação dos alunos com a disciplina.

Introdução - Objetivos

O objetivo deste trabalho é apresentar um modelo de seleção de características que melhore os resultados de sistemas de avaliação de questões discursivas curtas.

- Reproduzir os experimentos realizados por Pissinati (2014);
- Identificar os itens necessários para aproximação do modelo às correções do especialista;
- Desenvolver sistemas classificadores para comparação da efetividade da seleção de características;
- Verificar a validade da proposta conforme os resultados da avaliação;
- Discutir os resultados obtidos apresentando a eficácia dos modelos;
- Gerar um modelo de feedback que apresente ao usuários os modelos criados.

Avaliação Assistida por Computadores



O estudo realizado por Page (1968) é considerado o início da integração dos métodos computacionais para avaliação do aprendizado em questões discursivas.

Voltado para análise de redações, no final dos anos 90 as atividades foram interpretadas segundo os modelos de aprendizagem distintos para cada tipo ou categoria da resposta esperada como apresenta a figura proposta por Burrows et.al. (2015).

Figura1. Categorização feita por Burrows et.al. (2015) aos tipos de atividade segundo seu modelo de conhecimento.

Avaliação Assistida por Computadores

Então, cada atividade textual foi trabalhada de formas distintas e novas técnicas aplicadas à essa temática (Burrows et.al., 2015). Assim, a correção automática de redações mantém foco no estilo de escrita e as demais respostas discursivas buscam os conteúdos chave do critério avaliativo do professor.

Tipo de Atividade	Tamanho	Foco	Liberdade Textual
Preenchimento	De uma até poucas palavras	Palavra	Fixo
Respostas Curtas	De uma frase até um parágrafo	Conteúdo	Fechado
Redações	De dois parágrafos até algumas páginas	Estilo	Aberto

Tabela 1. Tradução da distinção feita por Burrows et.al. (2015) para os tipos de atividades textuais.

Avaliação Assistida por Computadores

A objetividade das questões influencia diretamente na capacidade de processamento de questões discursivas. Segundo Bailey e Meurers (2008) a capacidade de processamento dessas atividades e a aquisição de bons resultados depende da restrição de conteúdo.

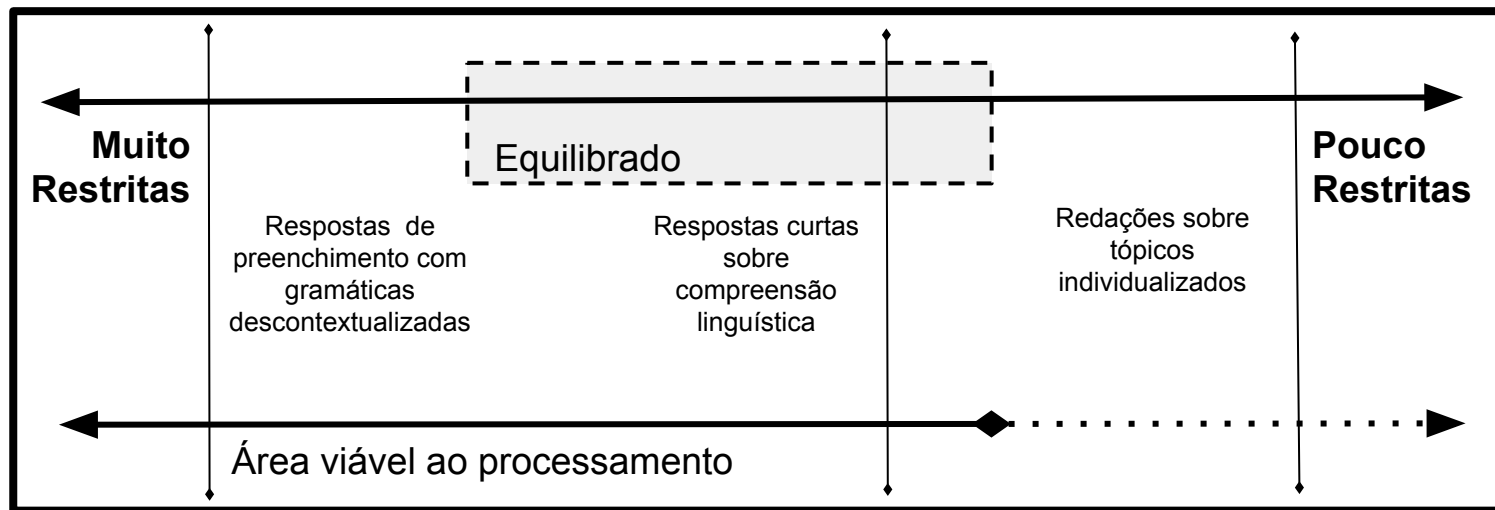


Figura 2. Capacidade de processamento para questões discursivas e a limitação dado ao escopo da atividade segundo Bailey e Meurers (2008).

Avaliação de Respostas Discursivas

1. Dividir as respostas em K grupos.
2. Selecionar para cada grupo os pares de maior similaridade e menor similaridade, e as respostas com maior e menor quantidade de termos para treinamento do algoritmo.
3. Corrigir as atividades selecionadas.
4. Utilizar as classes de nota para descoberta de padrões representativos.
5. Eliminar as demais características
6. Utilizar os padrões na classificação dos demais documentos
7. Retornar *feedbacks* e resultados aos usuários.

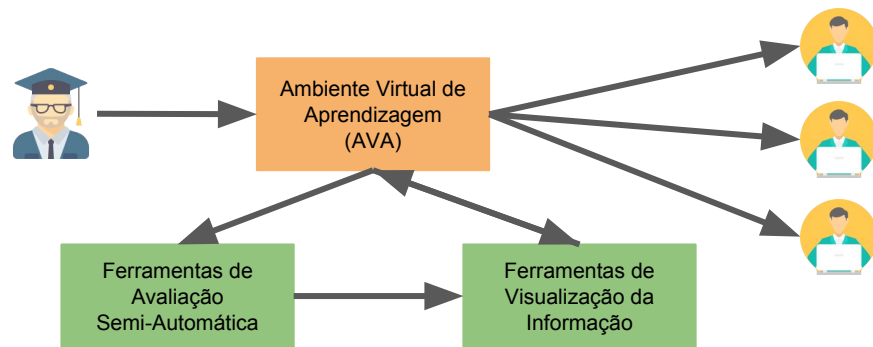


Figura 3. Modelo de interatividade entre sistemas, professores e alunos

Avaliação de Respostas Discursivas

O mapa de características é uma ferramenta de extração e visualização de informações para identificação dos principais elementos no texto do aluno com a finalidade de reduzir os esforços de verificação de aprendizagem e revisão do conteúdo. Tal intuito também pode ser encontrado no desenvolvimento de ferramentas como o *Willow* (Perez-Marin et. al., 2005), o Módulo *Rouge* (Gütl, 2007), *TreeMap* (Pissinati, 2014) e o *FreeText Author* (Jordan e Mitchell, 2009).

Feedback

Nota	100,00 / 100,00
Avaliado em	domingo, 12 Feb 2017, 11:08
Avaliado por	 Moodle WS
Comentários de feedback	 Exercício 5 Os elementos base no modelo proposto pelo DCC Digital Curation Center são Dados objetos digitais base de dados e também...

Figura 4. Mapa de características apresentado no AVA Moodle.

Avaliação de Respostas Discursivas

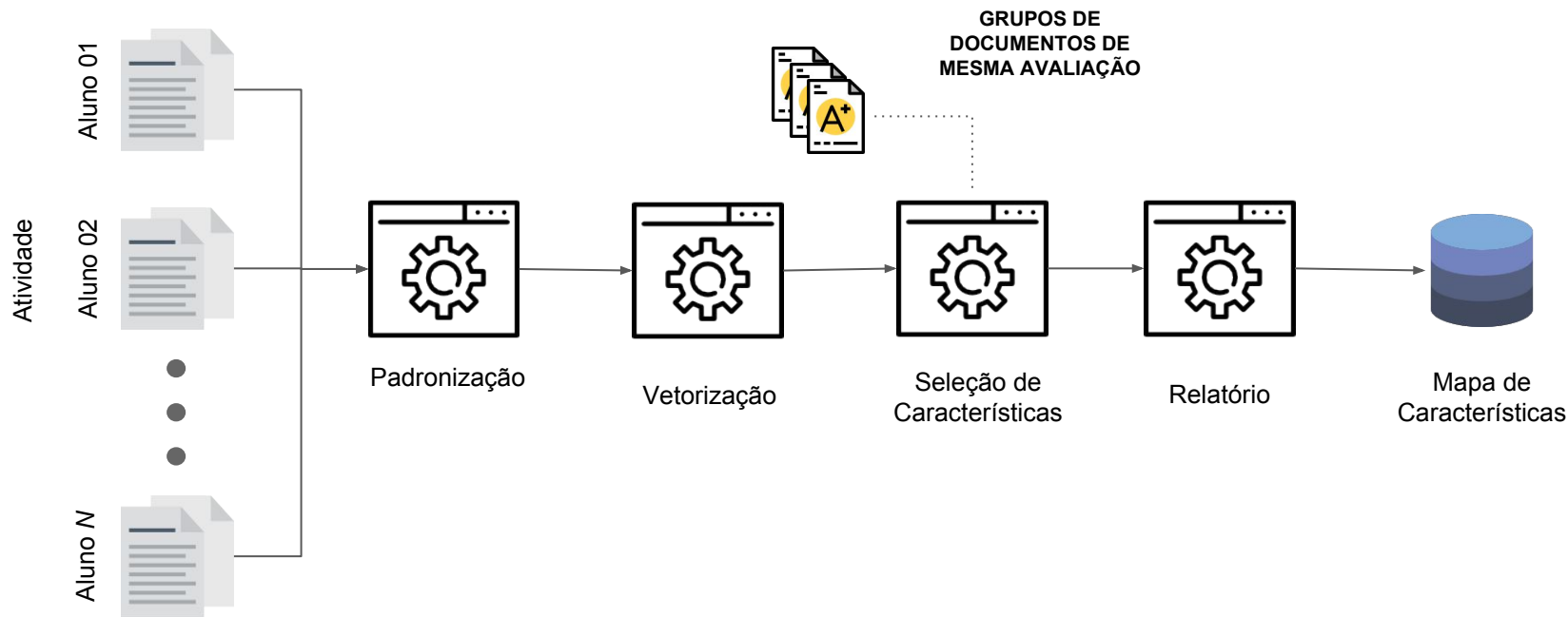


Figura 5. Processo de avaliação e construção do mapa de características após a avaliação de todos os itens de treino pelo professor.

Através de um **servidor WEB** que **irá armazenar** os **dados** para futuras interações

Criando campos ocultos que serão usados por scripts de processamento para **armazenar** informações

As informações ficam **armazenadas** em um **servidor web** e **através** de um motor de busca buscador é **possível** que estas informações seja consultadas pelo **usuário**

É **possível** obter informações de **usuários** em um documento eletrônico **através** de **formulários** a ser preenchido por ele um campo dentro do próprio documento

De varias maneiras como **formularios** com metodos GET e POST Sessões e Cookies

É **possível** devido ao **banco** de **dados**

Basta usar um **servidor** para **armazenar** as informações **enviadas** pelo **usuário** e recuperá **las** mais **tarde** para uso

É **possível** obter informações de leitor **usuário** **através** da **criação** de um **formulário** Existem duas partes no **formulário** a **coleção** de campo **rótulos** e botões e **script** de processamento

Quando o **usuário** **entra** numa **rede** precisa preencher **formulários** que são salvos e **enviados** aos **servidor** sendo assim as **preferências** dele **estarão** registradas para os **dados** das próximas navegações

Através da **computação** em nuvem

Uma das praticas é o **armazenamento** da informação **através** dos **dados** em _Formulários_ **HTML** como é feita a aplicação que **receberá** essas informações no _lado_ do **servidor**

Por meio do **formulário** que **possibilita** a **interação** do usuário e do **servidor** é **possível** **armazenar** as **preferencias** dos usuários no **banco** de **dados** **após** a **análise** do **Script**

É **possível** por meio de uma **servidor** **onde** o **site** esta **hospedado** que tem uma linguagem e uma folha de estilos e tenha ainda uma **banco** de **dados**

O sistema **operacional** **recebe** os **comandos** **enviados** pelo **usuário** Os sistemas operacionais **utilizam** interfaces **gráficas** para **facilitar** a **interação** do **usuário** muitos **comandos** podem ser executados **através** dos ícones

Atraves de um serviços de Assinatura Digital de Documentos Eletrônicos

Utilizando **interação** por **formulário** **formulários** estes ligados a **bancos** de **dados**

Sim é **possível** **utilizando** **tecnologias** de construção de **paginas** na **grande** **rede** de **computadores** **através** de **programação** com **html** por exemplo

Quando o **usuário** fornece suas informações em determinado lugar são gravadas e usadas como referência futuramente

Por meio de **formulários** preenchidos pelo leitor **usuário** ao acessar tais documentos desde newsletter ou Comment Form até **formulários** de cadastro se for o caso

repositórios de documentação científica podemos pensar na integração de **repositórios** **digitais** **Conhecer** as propriedades físicas dos suportes a serem **utilizados** para a migração Saber criar e manter metadados de preservação digital

LEGENDA

Nota 100.0 **Nota 90.0** **Nota 0.0**

Figura 5. Mapa de características para a atividade Tec-II-5-186

Avaliação de Respostas Discursivas

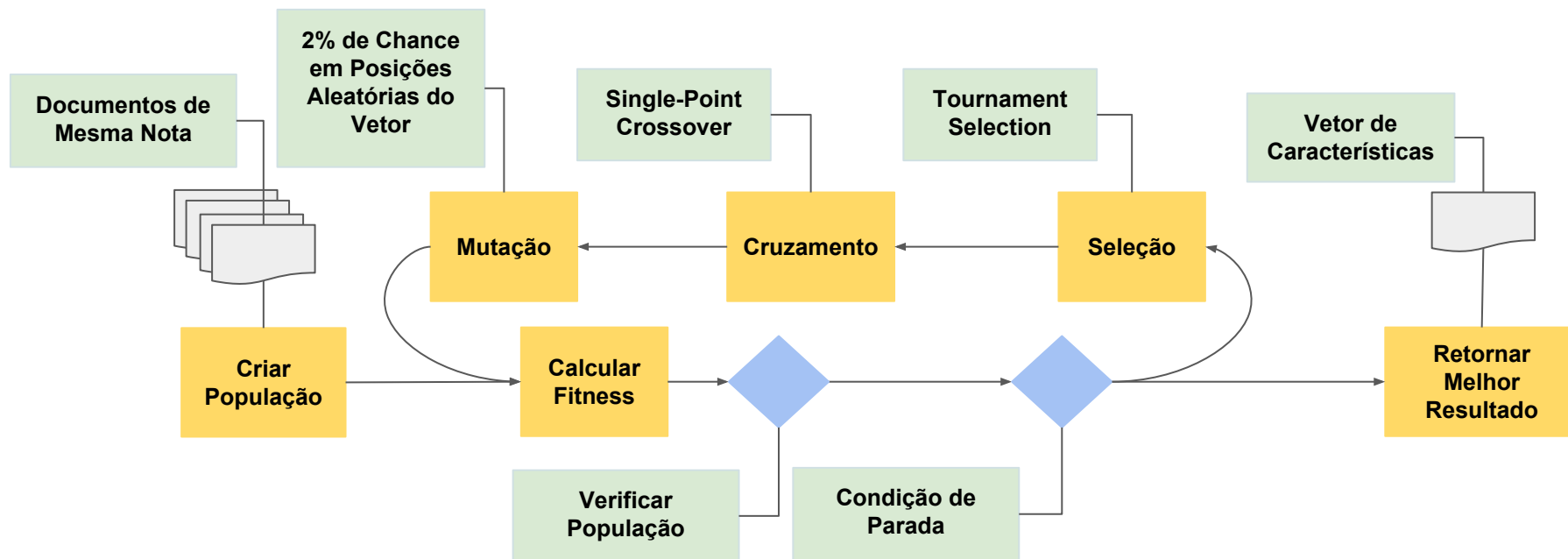


Figura 6. Estrutura do G.A. para seleção de características nas classes de nota.

Experimentos e Resultados

- Métricas para avaliação por classes:

- **Accuracy**

$$Accuracy(y_{real}, y_{pred}) = \frac{1}{n_{amostras}} \sum_{i=0}^{n_{amostras}-1} 1(y_{i_{real}} = y_{i_{pred}})$$

- **Precision**

$$Precision(y_{real}, y_{pred}) = \frac{T_P}{T_P + F_P}$$

- **Recall**

$$Recall(y_{real}, y_{pred}) = \frac{T_P}{T_P + F_N}$$

- **F-Score**

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$

- Métricas para avaliação contínua:

- **Erro Médio Absoluto - MAE**

$$MAE(y_{real}, y_{pred}) = \frac{1}{n_{amostras}} \sum_{i=0}^{n_{amostras}-1} |y_{i_{real}} - y_{i_{pred}}|$$

- **Desvio Padrão do Erro - SEM**

$$MSE(y_{real}, y_{pred}) = \frac{1}{n_{amostras}} \sum_{i=0}^{n_{amostras}-1} (y_{i_{real}} - y_{i_{pred}})^2$$

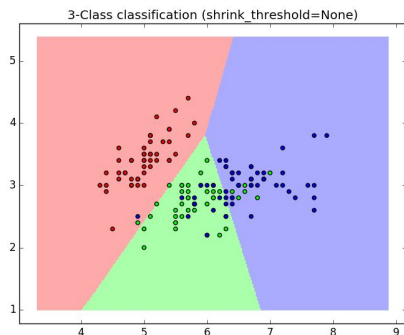
- **Erro Quadrático Médio - MSE**

$$SEM(y_{real}, y_{pred}) = \frac{\sigma}{\sqrt{n_{amostras}}}$$

Experimentos e Resultados

Centroid Based Classifier (CBC)

- Associação ao padrão médio da classe
- Classificação é dada pela influência de cada centroide de classe
- Visa remontar os *clusters* da etapa de treinamento



K-Nearest Neighbors (K-NN)

- Associação aos padrões de treinamento
- Classificação é dada diretamente pela influência de cada dado
- Visa salientar padrões únicos de avaliação

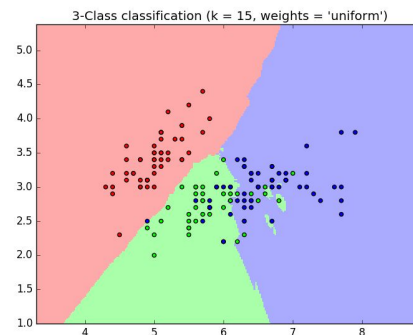
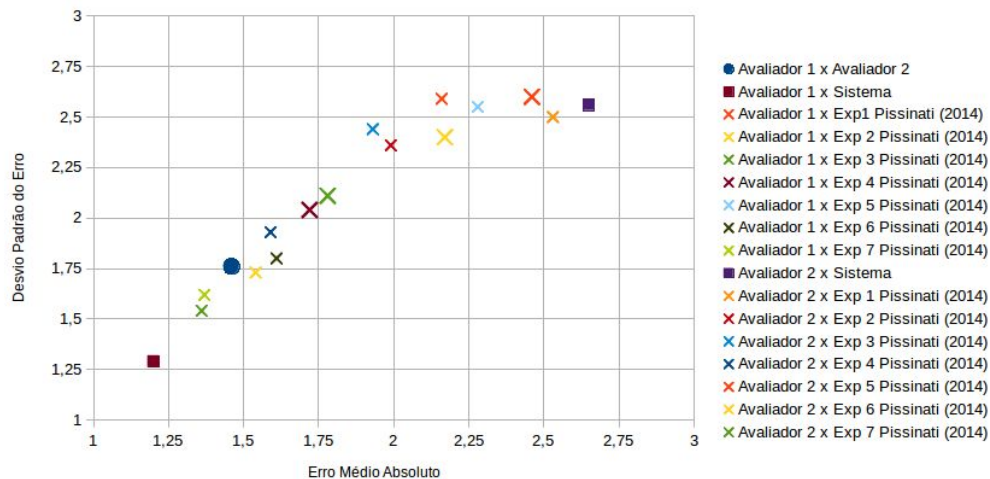


Figura 7 e 8. Exemplos da classificação bidimensional com o *dataset Iris* para respectivamente CBC e KNN.

Experimentos e Resultados

- **Vestibular UFES**
 - 5 atividades com 460 respostas transcritas do vestibular de 2012 da Universidade Federal do Espírito Santo - UFES.
- **Disciplinas UFES**
 - 45 atividades com coletadas via Moodle de disciplinas ministradas na Universidade Federal do Espírito Santo - UFES.
- **Texas Dataset**
 - 87 atividades com, em média, 30 respostas cada, da disciplina de Estrutura de Dados da Universidade do Norte do Texas.
- **Concurso da Hewlett Foundation no Kaggle**
 - 10 atividades com 17043 respostas disponibilizadas pelo ASAP-Kaggle.

Experimentos e Resultados - Vest UFES



Base de dados transcrita do vestibular de 2012 de respostas das questões discursivas de Língua Portuguesa.

Utilizada para comparação dos modelos de avaliação de dois avaliadores humanos e o corretor proposto por Pissinati (2014).

Esse autor utilizou 7 experimentos cumulativos segundo os resultados prévios. Os valores apresentados são relativos ao último experimento.

	Especialista		Sistema		Exp. 7 Pissinati (2014)	
	MAE	SEM	MAE	SEM	MAE	SEM
Corretor						
Avaliador 1	1,46	1,76	1,20	1,29	1,37	1,62
Avaliador 2	1,46	1,76	2,65	2,56	1,36	1,53

Tabela 2. Resultados de MAE e SEM para cada avaliador comparado com o especialista.

Experimentos e Resultados - Vest UFES

QUESTÃO 5: Com base nos elementos constitutivos do ato de comunicação, Roman Jakobson estabeleceu seis funções da linguagem (e a ênfase de cada uma delas): referencial (ênfase no assunto; no conteúdo), emotiva (ênfase no emissor; no sujeito), conativa (ênfase no receptor; no interlocutor), poética (ênfase na forma; na construção), metalinguística (ênfase no código; na autorreferência) e fática (ênfase no canal; no contato).

- "O navio negreiro" - Castro Alves
- "7" - Mário de Sá-Carneiro
- "Os arredores florem" - Paulo Roberto Sodré

Escolha um dos textos da 4 a Questão indique e explique a ocorrência de uma dessas funções.

Experimentos e Resultados - Disciplinas UFES

Base de dados composta por 45 atividades de diferentes disciplinas da UFES.

O resultado médio de *precision* e *recall* foi de 94,24% e 88,64%, respectivamente, para o K-NN e 92,86% e 80,00% para o CBC.

O número de características, em média, foi reduzido para 25% do total.

Segundo os autores são os **arquivistas** e **bibliotecários**

Os **arquivistas** e **bibliotecários**

Os **Arquivistas** e os **Bibliotecários**

Arquivista e **bibliotecário**

Segundo o autor os tradicionais **profissionais** da informação são os **arquivistas** e os **bibliotecários**

Conforme o texto Os **profissionais** da informação são **Arquivistas** e **Bibliotecários**

Arquivistas e **bibliotecários**

Arquivistas e **Bibliotecários**

Arquivista e **Bibliotecário**

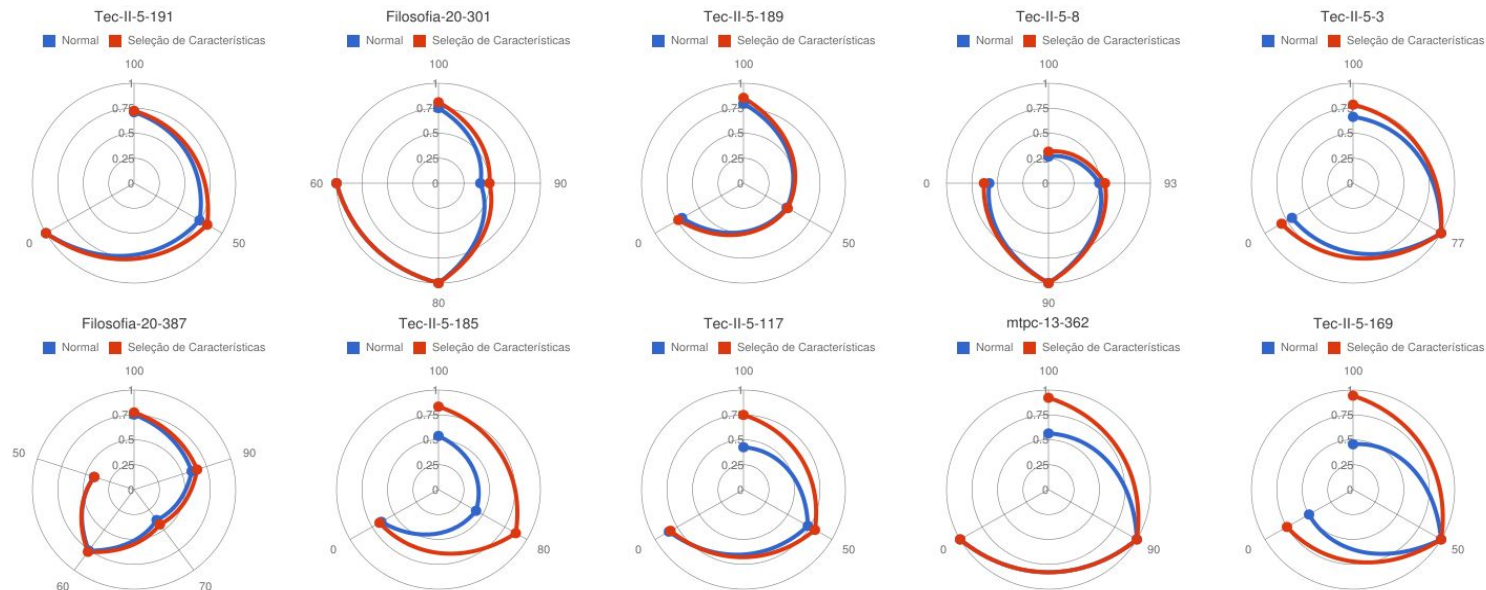
Arquivistas e **Bibliotecários**

Arquivistas e **bibliotecários**

LEGENDA: **NOTA 100.0**

Experimentos e Resultados - Disciplinas UFES

Análise da similaridade antes (azul) e depois (vermelho) da seleção de características em cada classe de nota das atividades da base de dados de disciplinas da UFES.



Figuras 10. Similaridade interna por classe de nota em 10 atividades de disciplinas da UFES antes e depois da seleção de características.

Questão TEC_II-5-170

Pergunta: Cite algumas linguagens para o lado do servidor

São aquelas linguagens que são reconhecidas executadas e interpretadas pelo próprio servidor e que se enviam ao cliente em um formato compreensível para ele **css php mysql**

css php e MySQL

Banco de **dados** como **MySQL PHP ASP**

Banco de **dados MySQL PHP e ASP**

PHP CSS e MySQL

PHP mysql java JavaScript

PHP ASP NET Oracle Perl Java Python

JAVA ASP PHP

PHP ASP E NET

Linguagens do lado do servidor aquelas que são reconhecidas executadas e interpretadas pelo próprio e que se enviam ao cliente num formato compreensível Ex **PHP JAVA PYTHON NET ASP ETC**

No domínio da rede as linguagens de lado servidor mais amplamente **utilizadas** para o **desenvolvimento** de páginas dinâmicas são o **ASP** Active Server **Pager PHP PERL e Java**

PHP JAVA SCRIPT SQL

PHP ASP JAVA RUBY

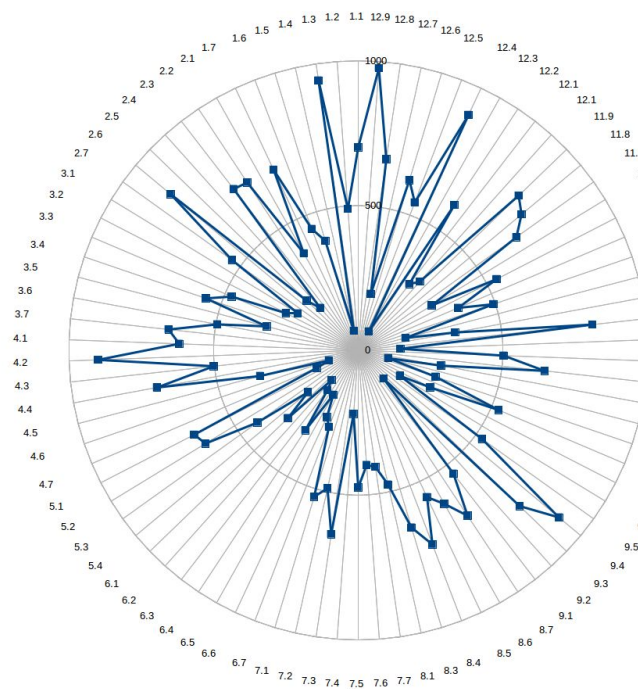
LEGENDA: **NOTA 100.0** **NOTA 90.0**

Experimentos e Resultados - Texas *Dataset*

Base de dados composta por 87 atividades da disciplina de Estrutura de Dados da Universidade do Norte do Texas.

O resultado médio de *precision* e *recall* foi de 87,85% e 81,80%, respectivamente, para o K-NN e 92,45% e 87,50% para o CBC.

O menor número de características do *dataset* foi de 69 e o maior de 977 termos.



Figuras 12. Número de características para cada atividade do *Texas Dataset*

Questão Texas Dataset 9.7

Pergunta: *What data structure is more appropriate for scheduling printing jobs at a printer, a stack or a queue?*

Chave de Resposta: *Queue*

Nota 1.0:

A stack some **printing jobs might** have a **higher priority** than **others** and they can be easily **inserted** at the front of the **list** or **anywhere** between

Nota 5.0:

a **queue**

Nota 5.0:

Queue s are more **appropriate** as **printers** are **usually first come first serve**

Nota 5.0:

Queue br

Nota 5.0:

A **Queue data structure** is more **appropriate** for **scheduling printing jobs** at a **printer** The **first job** sent to the **printer** is the **first job printed**

Nota 5.0:

a **queue** It wouldn t be **unfair** for the **first job** to **finish last**

Nota 5.0:

A **queue** is more **appropriate** because the clients which have **waited** the **longest** should be the **first served**

Nota 5.0:

Queue

Nota 5.0:

a **queue** because you **want** the **jobs** to **execute** in the **order** that they are **received**

Nota 5.0:

Queue its better to be **FIFO** than LIFO stack

Nota 5.0:

queue because it **guarantees** that **job** are **executed** in the **order** they **come in**

Nota 5.0:

Queues would be preferable to stacks for **use** as **scheduling print jobs** because it **would print jobs** in the **order** that they were **sent** to the **printer**

LEGENDA: **NOTA 5.0**

Figuras 13. Mapa de características para a atividade Texas Dataset 9.7.

Questão Texas Dataset 2.6

Pergunta: *What is the difference between a function prototype and a function definition?*

Chave de Resposta: *A function prototype includes the function signature, i.e., the name of the function, the return type, and the parameters' type. The function definition includes the actual body of the function.*

Nota 5.0

A **function definition** does not **require** any **additional information** that **needs** to be **passed inside** its **parenthesis** **br** to **execute** While a **definition** **prototype** **requires** more than **one parameters** to be **passed in order** to **complete** its **br** **task**

Nota 5.0

Function prototypes **describe** the **class s** **public interface**

Nota 5.0

A **function definition** **contains** all the **code** for a **function** to **work** A **function prototype** just **shows** the **output input** and **function name**

Nota 5.0

function prototype **describe** the **class s** **public interface** **without revealing** **br br** the **class s** **member function implementations** **function definitions** **show** what **br br** **implementations** are being **done**

Nota 5.0

A **function prototype** only **names** the **function** its **return type** and it s **argument list** while a **definition** **defines** the above as **well** as what the **function actually** does

Nota 5.0

A **function prototype** is a **declaration** of a **function** that **tells** the **compiler** the **function s** **name** its **return type** and the **types** of its **parameters**

LEGENDA: **NOTA 5.0** **NOTA 4.0** **NOTA 3.0**

Questão

Texas Dataset 1.5

Pergunta:

what is a variable?

Chave de Resposta:

a location in memory that can store a value

Nota 1.0

Variable can be a integer or a **string** in a **program**

Nota 2.0

A **named** object that can **hold** a numerical or letter **value**

Nota 3.0

An identifier that **holds** a **location** in **memory**

Nota 3.0

a **symbol** that **stands** in for a **value** that **may** or **may not change depending** on the **program**

Nota 4.0

a **stored value** used by the **program**

Nota 5.0

a block of **memory** that **holds** a specific type of **data**

Nota 5.0

It is a **location** in the **computer s memory** where it can be **stored** for **use** by a **program**

Nota 5.0

A **location** in **memory** where **value** can be **stored**

Nota 5.0

A **variable** is the **location** in a **computer s memory** where a **value** can be **stored** for **use** by a **program**

Nota 5.0

a **location** in **memory** where **data** can be **stored** and retrieved

Nota 5.0

Variable is a **location** in the **computer s memory** in which a **value** can be **stored** and later can retrieve that **value**

Nota 5.0

location in **memory** where a **value** can be **stored**

LEGENDA: **NOTA 5.0** **NOTA 4.0** **NOTA 3.0** **NOTA 2.0**

Figuras 15. Mapa de características para a atividade Texas Dataset 1.5.

Experimentos e Resultados - ASAG Kaggle

Base de dados da competição *Automatic Short Answer Grading* do Kaggle, com 10 atividades variando o conteúdo de artes à ciências. Apresenta alto índice de sobreposição (*overlapping*) e número muito superior de respostas quando comparada com as demais bases de dados.

Dataset Kaggle	1	2	3	4	5	6	7	8	9	10
Nº de Grupos	38	46	30	30	41	59	39	56	44	45
Treino (em %)	12,56	19,25	8,99	9,90	12,48	16,92	12,12	16,51	13,40	15,24
Treino (em qtd.)	210	246	170	172	224	304	218	297	241	250
Total de Respostas	1672	1278	1891	1738	1795	1797	1799	1799	1798	1640

Tabela 3. Distribuição de treino para cada uma das dez atividades do Kaggle.

Experimentos e Resultados - ASAG Kaggle

Precisão média de 74,43% para o Avaliador 1 e 75,74% para o Avaliador 2.

Em 7 resultados com precisão acima dos 70%, 3 deles acima dos 90%.

A seleção de características destaca-se nas atividades 7 e 8, onde induziu um aumento de aproximadamente 10% na precisão da classificação.

Dataset	Avaliador	Antes da Seleção de Características				Depois da Seleção de Características			
		KNN		CBC		KNN		CBC	
		Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
KAGGLE 1	1	0,5260	0,3768	0,5129	0,4839	0,5592	0,3451	0,5613	0,4605
	2	0,5319	0,3780	0,5176	0,4934	0,5587	0,3397	0,5787	0,5042
KAGGLE 2	1	0,4361	0,3529	0,4918	0,4898	0,3944	0,3850	0,5063	0,4867
	2	0,4306	0,3513	0,4847	0,4812	0,4746	0,4249	0,5184	0,4969
KAGGLE 3	1	0,5529	0,4823	0,4444	0,4469	0,5759	0,4014	0,4901	0,4617
	2	0,6063	0,5098	0,4980	0,4944	0,5865	0,4971	0,5334	0,5056
KAGGLE 4	1	0,8777	0,4931	0,6721	0,6358	0,9007	0,4810	0,7284	0,6133
	2	0,8890	0,4781	0,6253	0,6197	0,9043	0,4764	0,7024	0,5875
KAGGLE 5	1	0,9760	0,7989	0,8213	0,8123	0,9779	0,7972	0,8962	0,8290
	2	0,9762	0,7967	0,8315	0,8078	0,9749	0,7961	0,8858	0,8340
KAGGLE 6	1	0,9579	0,8731	0,7172	0,7323	0,9608	0,8809	0,8490	0,8548
	2	0,9593	0,8709	0,7293	0,7490	0,9720	0,8709	0,7888	0,7997
KAGGLE 7	1	0,7644	0,6459	0,6521	0,6165	0,5774	0,5264	0,7418	0,6348
	2	0,7630	0,6465	0,6515	0,6115	0,5936	0,5192	0,7190	0,6092
KAGGLE 8	1	0,6798	0,5431	0,5957	0,6009	0,7831	0,4247	0,6633	0,5264
	2	0,6182	0,5425	0,6101	0,6087	0,7569	0,4330	0,6070	0,5520
KAGGLE 9	1	0,7659	0,4149	0,6379	0,6307	0,7268	0,3849	0,6420	0,5512
	2	0,8016	0,4705	0,6410	0,6346	0,8497	0,3826	0,6624	0,5423
KAGGLE 10	1	0,6550	0,4537	0,6547	0,6415	0,7081	0,3909	0,6427	0,6311
	2	0,6591	0,4579	0,6432	0,6299	0,7133	0,4128	0,6074	0,5854

Tabela 4. Resultado de *precision* e *recall* para cada avaliador e cada classificador, antes e depois da seleção de características

Conclusões

As bases de dados para testes específicos de desempenho como o Kaggle e o Vest UFES mostram a competência do sistema diante dos desafios apresentados como a avaliação contínua e seleção de padrões em grandes bases de dados.

A precisão média acima de 90% nas bases de dados de disciplinas da UFES e Texas Dataset, podemos então dizer que o sistema reconhece adequadamente os modelos avaliativos e a identificação de critérios avaliativos é validada pelo mapa de características.

Trabalhos Futuros

- Aprimorar a seleção de características conforme a distribuição analisada na etapa de *clustering*.
- Inclusão de procedimentos de remoção de *outliers*
- A construção de relatórios que aproximem os resultados aos critérios de correção do professor. Inclusive através da criação do quadro de *rubrics* (ARTER; CHAPPUIS, 2006).

- ASSUNÇÃO, A. Á.; OLIVEIRA, D. A. Intensificação do Trabalho e Saúde dos Professores. *Educação & Sociedade*, scielo, v. 30, p. 349 – 372, 08 2009.
- LIMA, J. d. O. d.; GIRAFFA, L. M. M. Um Estudo Sobre Funcionalidades Dos Softwares Educacionais Para Suporte A Aprendizagem De Matemática Do Ensino Médio. Porto Alegre, RS, 4 2006.
- BAILEY, S.; MEURERS, D. Diagnosing Meaning Errors in Short Answers to Reading Comprehension Questions. In: *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications*, held at ACL 2008. Columbus, Ohio, USA: Association for Computational Linguistics, 2008. p. 107–115.
- BARREIRA, C.; BOAVIDA, J.; ARAÚJO, N. Avaliação Formativa: Novas Formas de Ensinar e Aprender. *Revista Portuguesa de Pedagogia*, Universidade de Coimbra, v. 40, n. 3, p. 95–133, 2006.
- BURROWS, S.; GUREVYCH, I.; STEIN, B. The Eras and Trends of Automatic Short Answer Grading. *International Journal of Artificial Intelligence in Education*, v. 25, n. 1, p. 60–117, 2015.
- JORDAN, Sally; MITCHELL, Tom. e-Assessment for learning? The potential of short-answer free-text questions with tailored feedback. *British Journal of Educational Technology*, v. 40, n. 2, p. 371-385, 2009.
- MARTENS, V. M. Nivelamento de Habilidades de Leitura e Escrita: Um Fazer Pedagógico nas Escolas Públicas de Ensino Integral Frente ao Fracasso Escolar. *Construção psicopedagógica*, scieloepsic, v. 24, p. 119–132, 2016.
- MOHLER, M.; MIHALCEA, R. Text-to-text semantic similarity for automatic short answer grading. *12th Conference of the European Chapter of the Association for Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009.
- OLIVEIRA, K. L. d.; SANTOS, A. A. A. Compreensão em Leitura e Avaliação da Aprendizagem em Universitários. *Psicologia: Reflexão e Crítica*, scielo, v. 18, p. 118 – 124, 04 2005.
- PEREZ-MARÍN, D. et al. Willow: Automatic and Adaptive Assessment of Students Free-Text Answers. *Sociedade Espanhola de Processamento de Linguagem Natural(SEPLN)*, v. 37, p. 367–368, 2005
- PAGE, E. B.; PAULUS, D. H. *The Analysis of Essays by Computer. Final Report*. Washington, DC: Office of Education (DHEW), 1968.
- PISSINATI, E. M. Uma Proposta de Correção Semi-Automática de Questões Discursivas e de Visualização de Atividades para Apoio à Atuação do Docente. Vitória, ES, 2014.

PPGI - Programa de Pós Graduação em Informática - Mestrado em Informática
Universidade Federal do Espírito Santo



OBRIGADO!

Marcos A. Spalenza

marcos.spalenza@gmail.com