A monografia de qualificação

Ivandré Paraboni

USP / EACH

Programa de Pós-graduação em Sistemas de Informação (PPgSI)

Produtos da disciplina

<>

Qualificação

O que deve ter sido obtido até aqui

- Uma lista de artigos de interesse (PDF + BibTeX)
- Estratégia geral para uma possível RS
- Os 10 itens de projeto

Como usar isso na qualificação

- Uma lista de artigos de interesse (PDF + BibTeX)
- Estratégia geral para uma possível RS
- Os 10 itens de projeto

Ler e escrever sobre os trabalhos selecionados;

Conduzir a RS completa (e possivelmente escrever artigo)

Como usar isso na qualificação

- Uma lista de artigos de interesse (PDF + BibTeX)
- Estratégia geral para uma possível RS
- Os 10 itens de projeto

Conteúdo do capítulo da proposta de pesquisa;

Subsídios para a introdução da monografia

Ler e escrever sobre os trabalhos selecionados;

Conduzir a RS completa (e possivelmente escrever artigo)

Estrutura típica em capítulos

(1) Introdução
(2) Conceitos fundamentais
(3) Revisão bibliográfica
(4) Estudo piloto
(5) Proposta de pesquisa

} Estudos de terceiros

Ordem de escrita sugerida

(3) Revisão bibliográfica

(4) Estudo piloto

(5) Proposta de pesquisa

(2) Conceitos fundamentais

(1) Introdução

Por que????

Dependências entre conteúdos

Ordem de escrita sugerida

(3) Revisão bibliográfica

traz motivação p/proposta

é o mais difícil e mais demorado

não há qualificação sem revisão bibliográfica

Ordem de escrita sugerida

(3) Revisão bibliográfica

traz motivação p/proposta

é o mais difícil e mais demorado

não há qualificação sem revisão bibliográfica

(4) Estudo piloto

traz motivação adicional e outros elementos p/proposta

Ordem de escrita sugerida

(3) Revisão bibliográfica

traz motivação p/proposta

é o mais difícil e mais demorado

não há qualificação sem revisão bibliográfica

(4) Estudo piloto

traz motivação adicional e outros elementos p/proposta

(5) Proposta de pesquisa

conteúdo principal do documento

Parcialmente motivada pela revisão e resultados preliminares (cap.3,4)

Ordem de escrita sugerida

(3) Revisão bibliográfica

traz motivação p/proposta

é o mais difícil e mais demorado

não há qualificação sem revisão bibliográfica

(4) Estudo piloto

traz motivação adicional e outros elementos p/proposta

(5) Proposta de pesquisa

conteúdo principal do documento

Parcialmente motivada pela revisão e resultados preliminares (cap.3,4)

(2) Conceitos fundamentais

explica todos os termos, conceitos etc. usados nos outros capítulos

Ordem de escrita sugerida

(3) Revisão bibliográfica

traz motivação p/proposta

é o mais difícil e mais demorado

não há qualificação sem revisão bibliográfica

(4) Estudo piloto

traz motivação adicional e outros elementos p/proposta

(5) Proposta de pesquisa

conteúdo principal do documento

Parcialmente motivada pela revisão e resultados preliminares (cap.3,4)

(2) Conceitos fundamentais

explica todos os termos, conceitos etc. usados nos outros capítulos

(1) Introdução

Meta-descrição do documento inteiro

só pode ser redigida com clareza depois do resto todo

Agora na ordem 1..5

Cap. 1. Introdução

- Meta-descrição do documento como um todo
- É o último capítulo a ser escrito
  - Estratégia útil também para a escrita da introdução de artigos etc.

Cap. 1. Introdução

- Meta-descrição do documento como um todo
- É o último capítulo a ser escrito
  - Estratégia útil também para a escrita da introdução de artigos etc.
- Estrutura em grande parte semelhante à própria proposta
  - Mas em formato dissertativo (com poucos tópicos)
  - Tema > motivação > Lacuna > Hipótese > Objetivo etc.
  - Mas mais “alto nível” e resumido – sem excesso de detalhes

Cap. 1. Introdução

- Meta-descrição do documento como um todo
- É o último capítulo a ser escrito
  - Estratégia útil também para a escrita da introdução de artigos etc.
- Estrutura em grande parte semelhante à própria proposta
  - Mas em formato dissertativo (com poucos tópicos)
  - Tema > motivação > Lacuna > Hipótese > Objetivo etc.
  - Mas mais "alto nível" e resumido – sem excesso de detalhes
- Objetivo geral deve aparecer logo nas primeiras páginas, e com destaque em relação ao resto do texto

Cap. 2. Conceitos básicos

- Explica todos os recursos, técnicas, modelos etc. que são utilizados na sua pesquisa, ou de referência recorrente nos trabalhos relacionados
  - Se for necessário ao seu projeto, o nível de detalhamento deve ser maior

Cap. 2. Conceitos básicos

- Explica todos os recursos, técnicas, modelos etc. que são utilizados na sua pesquisa, ou de referência recorrente nos trabalhos relacionados
  - Se for necessário ao seu projeto, o nível de detalhamento deve ser maior
- Normalmente baseado em poucas fontes (artigos ou livros) de referência sobre o assunto
  - Não requer revisão sistemática ou de alguma forma exaustiva

Cap. 2. Conceitos básicos

- Explica todos os recursos, técnicas, modelos etc. que são utilizados na sua pesquisa, ou de referência recorrente nos trabalhos relacionados
  - Se for necessário ao seu projeto, o nível de detalhamento deve ser maior
- Normalmente baseado em poucas fontes (artigos ou livros) de referência sobre o assunto
  - Não requer revisão sistemática ou de alguma forma exaustiva
- Redação se torna mais produtiva se deixada para o final (ou pelo menos para depois da proposta), mantendo-se lista de todos os conceitos importantes a tratar

Cap. 2. Conceitos básicos

- Explica todos os recursos, técnicas, modelos etc. que são utilizados na sua pesquisa, ou de referência recorrente nos trabalhos relacionados
  - Se for necessário ao seu projeto, o nível de detalhamento deve ser maior
- Normalmente baseado em poucas fontes (artigos ou livros) de referência sobre o assunto
  - Não requer revisão sistemática ou de alguma forma exaustiva
- Redação se torna mais produtiva se deixada para o final (ou pelo menos para depois da proposta), mantendo-se lista de todos os conceitos importantes a tratar
- Não costuma ser um capítulo muito extenso

Cap. 3. Revisão Bibliográfica

- Pode ser exploratória ou sistemática
  - Mas deixe isso claro desde o início

Cap. 3. Revisão Bibliográfica

- Pode ser exploratória ou sistemática
  - Mas deixe isso claro desde o início
- Descrição do estado da arte
  - Trabalhos relacionados

Cap. 3. Revisão Bibliográfica

- Pode ser exploratória ou sistemática
  - Mas deixe isso claro desde o início
- Descrição do estado da arte
  - Trabalhos relacionados
- Em certos casos pode ser agregada ao capítulo 2

Cap. 3. Revisão Bibliográfica

- Pode ser exploratória ou sistemática
  - Mas deixe isso claro desde o início
- Descrição do estado da arte
  - Trabalhos relacionados
- Em certos casos pode ser agregada ao capítulo 2
- Normalmente assume a forma de uma série de resumos de trabalhos (ou grupos de trabalhos) relacionados
  - Seguindo algum tipo de estrutura hierárquica (ou no mínimo cronológica)

Cap. 3. Revisão Bibliográfica

- Pode ser exploratória ou sistemática
  - Mas deixe isso claro desde o início
- Descrição do estado da arte
  - Trabalhos relacionados
- Em certos casos pode ser agregada ao capítulo 2
- Normalmente assume a forma de uma série de resumos de trabalhos (ou grupos de trabalhos) relacionados
  - Seguindo algum tipo de estrutura hierárquica (ou no mínimo cronológica)
- O capítulo é finalizado com uma seção que resume o que foi estudado
  - Idealmente com algum tipo de quadro geral

Cap. 4. Estudo piloto (se houver)

- Qualquer atividade prática realizada durante o período até a qualificação, e de interesse para a proposta

Cap. 4. Estudo piloto (se houver)

- Qualquer atividade prática realizada durante o período até a qualificação, e de interesse para a proposta
- Pode assumir a forma de experimentos, implementação, coleta ou organização de conjuntos de dados, elaboração de questionários, entrevistas etc.
  - Recursos construídos (conjuntos de dados, *baselines*, ferramentas) a serem utilizados na proposta
  - Resultados parciais, ainda que negativos (motivação?)

Cap. 4. Estudo piloto (se houver)

- Qualquer atividade prática realizada durante o período até a qualificação, e de interesse para a proposta
- Pode assumir a forma de experimentos, implementação, coleta ou organização de conjuntos de dados, elaboração de questionários, entrevistas etc.
  - Recursos construídos (conjuntos de dados, *baselines*, ferramentas) a serem utilizados na proposta
  - Resultados parciais, ainda que negativos (motivação?)
- Serve também para demonstrar familiaridade com o tema e reforçar a viabilidade da proposta

Cap. 5. Proposta

- Os 10 itens desenvolvidos nesta disciplina
- Objetivos e hipóteses normalmente já foram enunciados no cap.1, e podem ser retomados aqui para maior clareza
  - Mas neste caso copie literalmente o texto usado anteriormente

Exemplo – Rafael F. S. Dias (2018)
(versão não corrigida)

Estrutura geral

- Divisão típica em 5 capítulos
- 53 páginas de conteúdo (cap. 1 ao 5)

Resultados parciais podem ser publicáveis

4 páginas!

1 Introdução

A caracterização autoral (CA) (do inglês, *Author Profiling*) é uma técnica computacional de reconhecimento de características de autores de textos com base em padrões linguísticos. Estes padrões são características baseadas em dialetos sociais que representam determinados grupos de pessoas que compartilham características sociais semelhantes, como a idade ou gênero (NGUYEN et al., 2016).

O surgimento de redes sociais fornece novos meios de comunicação e interações sociais. Com a geração de grandes volumes de dados textuais, apresenta a possibilidade de conhecer características de autores com base no que eles compartilham, não só para a predição de idade e gênero, como também para a predição de língua nativa, renda, escolaridade, ocupação etc., tornando a pesquisa de CA um campo amplo e de crescente interesse.

1.1 Objetivo

O objetivo geral deste trabalho é desenvolver modelos de aprendizado de máquina (AM) supervisionada baseados em RNAs e *word embeddings* para o reconhecimento de múltiplas tarefas de CA utilizando córpus de diversos domínios e idiomas, de modo que apresentem resultados superiores ao uso de modelos tradicionais.

De forma mais específica, consideramos desenvolver modelos de CA para reconhecimento de gênero, idade, localidade, nível de religiosidade, escolaridade, ocupação, entre outros, com base em textos rotulados provenientes de diversos córpus para os idiomas português, inglês e espanhol.

3 Revisão Bibliográfica

Este capítulo apresenta uma revisão bibliográfica exploratória sobre estudos de CA a partir de textos. Os estudos aqui considerados foram selecionados a partir dos repositórios ACL Anthology, Scopus, IEEE e PAN-CLEF, e foram privilegiados artigos mais recentes e de maior fator de impacto. Para facilitar a discussão, esta revisão é dividida em estudos que usam abordagens tradicionais de aprendizado de máquina (Seção 3.1) e estudos baseados em métodos de aprendizado profundo (Seção 3.2). Além disso, é apresentado um resumo dos estudos e considerações finais (Seção 3.3) sobre esta revisão.

3.1 Abordagens tradicionais de aprendizado de máquina

Tradicionalmente, modelos de CA fazem uso de métodos de regressão logística, árvores aleatórios e máquina de vetores de suporte (SVM), combinados a métodos de

3.3 Considerações

Este capítulo apresentou uma revisão bibliográfica acerca dos estudos de CA. O Quadro 1 apresenta uma visão geral desta revisão, contendo colunas com a citação de cada artigo, o conjunto de dados, o idioma, a tarefa de CA, o tipo de conhecimento e o método de aprendizado de máquina (AM) utilizado pelos estudos.

Os conjuntos de dados usados pelos estudos são, em sua maioria, da competição PAN-CLEF, entre as edições de 2013 e 2017, além de conjuntos próprios coletados do Twitter e de Blogs, conjuntos de dados públicos baseados no Facebook, assim como reviews de hotéis, livros da literatura do século XX e Kaggle.

Os idiomas considerados pelos estudos são inglês (EN), espanhol (ES), português (PT), italiano (IT), holandês (HO) e árabe (AR). Os estudos são, quase em sua totalidade, dedicados ao idioma inglês, e encontramos apenas um exclusivo para o idioma português (Guimarães et al., 2017).

Quadro 1 – Resumo dos trabalhos correlatos

Estudo	Dados	Idioma	Tarefas	Conhecimento	Método
Weren et al. (2014)	PAN-CLEF 2013	EN, ES	G, I	TF, IR	J48
Mechti, Jaoua e Belguith (2013)	PAN-CLEF 2013	EN, ES	G, I	TF	J48
Flekova, Preotiuc-Pietro e Ungar (2016)	Twitter	EN	I, R	Sintaxe	SVM
Meina et al. (2013)	PAN-CLEF 2013	EN, ES	G, I	TF, LSA	RandForest
Carmona et al. (2015)	PAN-CLEF 2015	EN, ES, IT, HO	G, I, P	LSA, SOA	SVM
López-Monroy et al. (2014)	PAN-CLEF 2014 e 2013	EN, ES, PT, AR	G, I	TF, SOA	LibLinear
Vollenbroek et al. (2016)	PAN-CLEF 2016	EN, ES, HO	G, I	word n-grams	SVM
Basile et al. (2017)	PAN-CLEF 2017 e 2016	EN, ES, PT, AR	I, V	word n-grams e char n-grams	SVM
Fatima et al. (2017)	RUEN-AP-2017	EN, AR	G, I	word n-grams e char n-grams	SVM
González-Gallardo et al. (2015)	PAN-CLEF 2015	EN, ES, IT, HO	G, I, P	char n-gramas e POS n-grams	SVM
Reddy, Vardhan e Reddy (2017)	TripAdvisor	EN	G	POS n-grams, TF-IDF	RegLog
Martinc et al. (2017)	PAN-CLEF 2017	EN, ES, PT, AR	G, V	POS n-grams	RegLog
Sap et al. (2014)	MyPersonality	EN	G, I	Léxico	SVM
Isbister, Kaati e Cohen (2017)	Blogs	EN, ES, FR, RU	G	LIWC	SVM
Guimarães et al. (2017)	Twitter	PT	I	*Word embeddings*	CNN
Sierra et al. (2017)	PAN-CLEF 2017	EN, ES, PT, AR	G, V	*Word embeddings*	CNN
Gopinathan e Berg (2017)	PAN[11] e Kaggle	EN	G	*Word embeddings*	CNN, LSTM
Bartle e Zheng (2015)	Blogs e Livros	EN	G	RCNN	RCNN
Kim et al. (2017)	Twitter	EN	G, I	RNN	LSTM

Fonte: Rafael Sandroni Dias (2018)

Exame de qualificação

Considerações

- Objetiva avaliar a proposta e a capacitação do candidato ao mestrado/doutorado
- Resultado favorável indica que o projeto é viável para a modalidade pretendida
- Quanto mais difícil, melhor! ☺

Obrigado