CISION

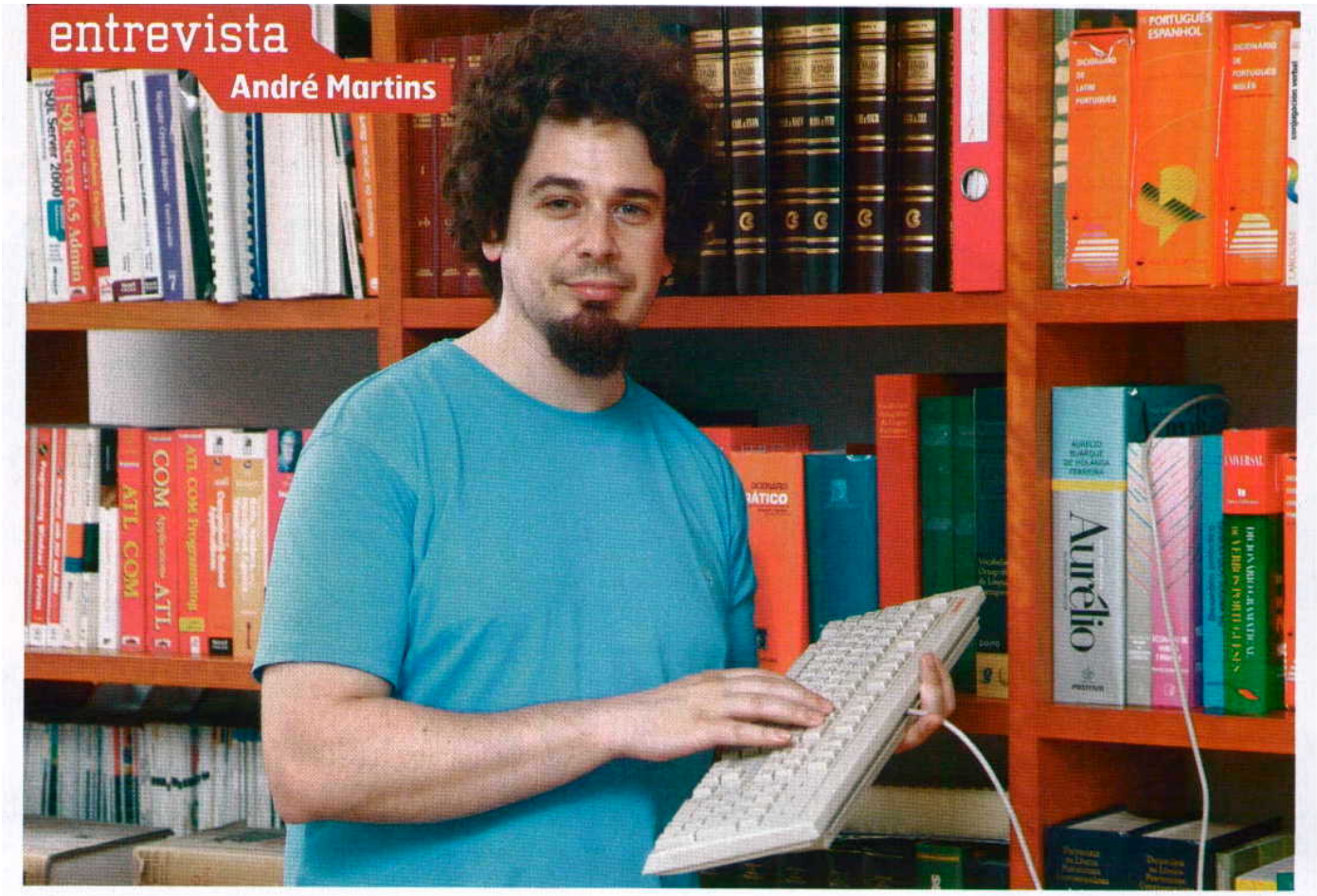
Exame Informática

ID: 43028545

01-08-2012

Tiragem: 20500
País: Portugal
Period.: Mensal
Âmbito: Tecnologias de Infor.

Pág: 26
Cores: Cor
Área: 19,13 x 28,06 cm²
Corte: 1 de 2



No idioma em que as máquinas falam

Os computadores não vão à escola para aprender a ler, mas com uma boa dose de estatística e Internet conseguem conversar com humanos

André Martins, investigador da Priberam, ganhou o Prémio Científico IBM de 2011 com o desenvolvimento de um modelo que usa a estatística para revolucionar a forma como os computadores percebem a linguagem dos humanos. O trabalho foi aplicado com sucesso em 14 idiomas.

Quanto tempo temos de esperar até um computador ter autonomia para responder a questões dos humanos?

Em alguns casos, o computador já é mais eficaz que as pessoas a organizar informação ou a usar de sistemas de resposta automática a perguntas. Um exemplo: a Priberam tem um sistema que permite colocar questões em linguagem natural. Esses sistemas conseguem processar um grande conjunto de documentos, que podem estar na Web, e dar uma resposta exata. Os computadores estão à nossa frente nesse campo. Mais complicada é a interpretação de texto, que permite que um computador perceba o significado de uma frase em linguagem natural. A minha tese de doutoramento e o trabalho premiado pela IBM têm esse objetivo.

Quais as principais dificuldades de comunicação de um computador?

É muito difícil lidar automaticamente com as negações, porque muitas vezes não se percebe qual é o objeto da negação. Também é difícil lidar com afirmações que apenas refletem uma possibilidade de algo acontecer, porque não podem ser mapeadas numa lógica de verdadeiro ou falso. Há problemas que não estão resolvidos e que estão relacionados com a ambiguidade... Num dos meus trabalhos tinha esta frase: «O Joaquim resolveu o problema com o método estatístico». Há duas interpretações possíveis nesta frase: 1) há um problema com o método estatístico e o Joaquim resolveu-o; ou 2) o Joaquim tinha um problema e usou o método estatístico para o resolver. Para um humano, é mais ou menos claro e automático, se tivermos informação de contexto... Mas fazer com que um computador consiga resolver esta ambiguidade é um problema...

... e curiosamente é com um método estatístico que o seu trabalho resolve os problemas causados pela ambiguidade.

A ambiguidade não existe só em textos literários. Trata-se de algo muito básico que está presente até num livro de matemática. Podemos tentar obter todas interpretações possíveis de uma frase - só que vamos deparar com uma explosão combinatória e torna-se complicado ter todas as

Foto: Pedro Melim

CISION

Exame Informática

ID: 43028545

01-08-2012

Tiragem: 20500
País: Portugal
Period.: Mensal
Âmbito: Tecnologias de Infor.

Pág: 27
Cores: Cor
Área: 18,32 x 25,77 cm²
Corte: 2 de 2

> "EM VEZ DE PESSOAS QUE ESCREVEM REGRAS TEREMOS MÁQUINAS QUE INFEREM REGRAS A PARTIR DE GRANDES VOLUMES DE TEXTO"

hipóteses em aberto. A estatística dá-nos essa ferramenta para resolver problemas, usando contextos observados numa fase em que o modelo é treinado. Assumimos que há um conjunto de documentos e assumimos que há anotações nesses documentos. Com base neste conjunto, há uma série de ferramentas de aprendizagem estatística que permitem fazer generalizações para frases que não têm anotações. Depois de treinar o modelo com 50 mil frases, tornou-se possível aplicá-lo numa frase nova para que possa tomar uma decisão (quanto ao significado).

A análise de uma estrutura sintática não seria suficiente para mudar a forma como usamos um sistema operativo?

Desde o início que a inteligência artificial tem por objetivo pôr os computadores a falar com os humanos. Em vez de executar um comando, passaríamos a dizer "abre aquele ficheiro". O meu trabalho pode ser visto como um módulo de uma coisa maior. Se se acoplar com um sistema de reconhecimento de voz pode dar origem a algo mais complexo. Nos anos 60, as pessoas que trabalhavam nesta área eram muito otimistas quando faziam previsões sobre robôs e outras coisas que acabaram por nunca acontecer. Atualmente, há sistemas que funcionam bem em domínios fechados. Por exemplo, os sistemas de atendimento de telefonemas. Mas não há um sistema que funcione em domínios abertos e aprenda com os dados que recebe.

Os portáteis de hoje têm potência computacional para correr modelos similares àquele que criou?

Hoje, a tendência é usar a computação paralela e haver cada vez mais processamento na nuvem e não nas máquinas que temos à frente. Tudo isto permite processar grandes quantidades de informação. Esse é que é grande paradigma novo. Durante o meu trabalho, conseguia analisar mais de 50 palavras, com um computador normal e em menos de um segundo.

Esse computador conseguiria ler, por exemplo, Os Maias?

Sim, seria muito rápido. Em pouco mais de um minuto, seria possível fazer a análise sintática desse texto.

E conseguiria saber que se trata de um romance proibido entre dois irmãos?

Conseguiria resolver ambiguidades como as que referi no exemplo que dei há pouco. A análise semântica ou literária será o próximo passo do que pretendo fazer. Está nos meus planos trabalhar neste tipo de coisas, porque são muito relevantes. Não tenho tanto interesse em textos literários... tenho interesse em textos jornalísticos, que são bastante objetivos, e em textos produzidos pelos próprios utilizadores. Podem ser blogues, tweets, redes sociais... Se conseguirmos processar esses textos, ficamos em condições de saber, por exemplo, qual o sentimento dominante face a um produto ou o assunto que é mais falado pelas pessoas.

... Os computadores não têm fome, não pagam IRS nem se chateiam com a namorada... Como é que eles podem realmente vir a escrever de forma equiparável à dos humanos?

Ainda não temos computadores como o HAL do "2001, Odisseia no Espaço". Os computadores ainda não conseguem tomar decisões de forma completamente autónoma. Nós não aprendemos a linguagem de forma totalmente isolada de todos os outros conhecimentos que adquirimos. Há uma conjugação da visão e da aquisição da linguagem... e essa conjugação supera cada uma das partes. Matematicamente não se conseguiu demonstrar que é possível uma aprendizagem automática que integra todas as fontes de informação. Quando isso acontecer passamos a ter computadores autónomos...

... e mais humanos!?

Sim. Há um projeto na Universidade de Carnegie Mellon que tem por base um computador que, 24 horas por dia, recolhe informação da Net. O objetivo é que o computador, à medida que faz a recolha, coloque informação em bases de dados, tome decisões, faça erros, e corrija os erros. Este tipo de coisas é mais realista. Mas ainda não está resolvido este desafio.

A Net pode ser a escola das máquinas?

Há 10 ou 15 anos chegámos a um novo paradigma, que está relacionado com a forma como se tira partido de uma grande quantidade de dados, que podem ser ruidosos, não tratados, cheios de erros, com informação falsa, páginas criadas por hackers, e spam... mas como a informação é tanta, permite extrair inteligência e aprender automaticamente algo.

O antropólogo do futuro será um computador?

Há um sinal muito forte na Internet e há também muito ruído. E interessa saber construir modelos e algoritmos que permitem distinguir o sinal do ruído. Nos próximos 10 a 20 anos, os grandes avanços vão assentar neste paradigma. É possível extrair informação sociológica, sociolinguística, e também se consegue extrair informação lexical, com conjuntos de dados muito grandes, que superam o que se pode produzir de forma manual, que é uma técnica clássica da inteligência artificial e que hoje está a ser substituída por este processamento de dados em massa. Em vez de termos pessoas que escrevem regras passamos a ter máquinas que conseguem inferir regras a partir de grandes volumes de texto.

Hugo Séneca

PERFIL

Com 34 anos de idade, André Martins é licenciado em Engenharia Eletrotécnica de Computadores e concluiu recentemente um doutoramento no âmbito da parceria entre o Estado Português e a Universidade de Carnegie Mellon University (CMU). **ANTES DA LINGUAGEM NATURAL, TRABALHOU NA ÁREA DA VISÃO POR COMPUTADOR.** Atualmente é investigador na Priberam.