



IA Aplicada à Saúde

UNIDADE 06

Processamento de Linguagem Natural Aplicado à Saúde

| Processamento de Linguagem Natural Aplicado à Saúde

A seguir alguns materiais os quais você pode revisitar o tema de PLN caso ache necessário.



SAIBA MAIS

Este livro foi disponibilizado de maneira gratuita no site da Universidade de Stanford, e está em constante atualização. Provavelmente é a maior referência do assunto disponível.

[Speech and Language Processing](#) por Dan Jurafsky and James H. Martin:

O Processamento de Linguagem Natural (PLN), juntamente às áreas já estudadas, compõe a relação de técnicas essenciais para IA em Saúde, isso porque ela é responsável por extrair dados das narrativas clínicas, que são consideradas as principais fontes de informação acerca do histórico de um paciente. Portanto, a grande maioria das soluções que necessitam destes dados, dependem das técnicas de PLN.

Mas antes de continuarmos, você lembra o que é PLN? PLN é um conjunto de técnicas utilizadas para fazer com que a máquina processe, manipule e interprete dados textuais. Você pode revisitar o tema nos links sugeridos a seguir.

OK, mas quais tipos de aplicações o PLN pode fomentar na área de saúde? A seguir algumas das principais possibilidades.

- **Melhoria do cuidado ao paciente – Apoio à decisão clínica**

1. Predição de complicações pós-cirúrgicas
2. Alerta de progressão de doenças
3. Sumarização de dados do paciente
4. Prevenção de readmissão do paciente
5. Gerenciamento do paciente oncológico ou crônico
6. Ferramenta de busca para medicina baseada em evidências

- **Redução de custos**

1. Cobrança eficiente: Atribuição de CID automática a partir da leitura do RES
2. Controle de erros administrativos: Evitar repetição de processos e atrasos em procedimentos

- **Eficiência administrativa**

1. Chatbots para redução de volume de atendimentos e triagem
2. Reconhecimento de fala para agilizar documentação
3. Rastreamento de pacientes, medicações e procedimentos para provisionamento de recursos

- **Apoio a indústria e pesquisa**

1. Recrutamento de pacientes para ensaios clínicos



SAIBA MAIS

Nos artigos científicos a seguir você pode ter uma noção geral mais detalhada do que pode ser feito com PLN para saúde.

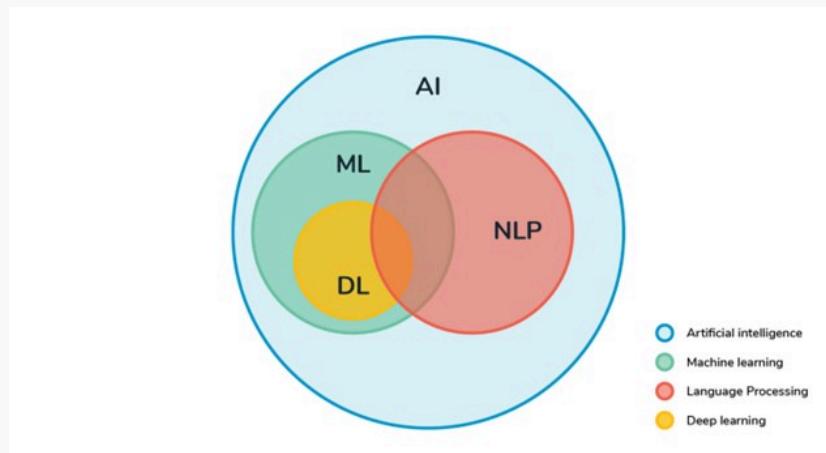
[The Revival of the Notes Field: Leveraging the Unstructured Content in Electronic Health Records](#)

[Clinical Text Data in Machine Learning: Systematic Review](#)

[Artificial intelligence approaches using natural language processing to advance EHR-based clinical research](#)

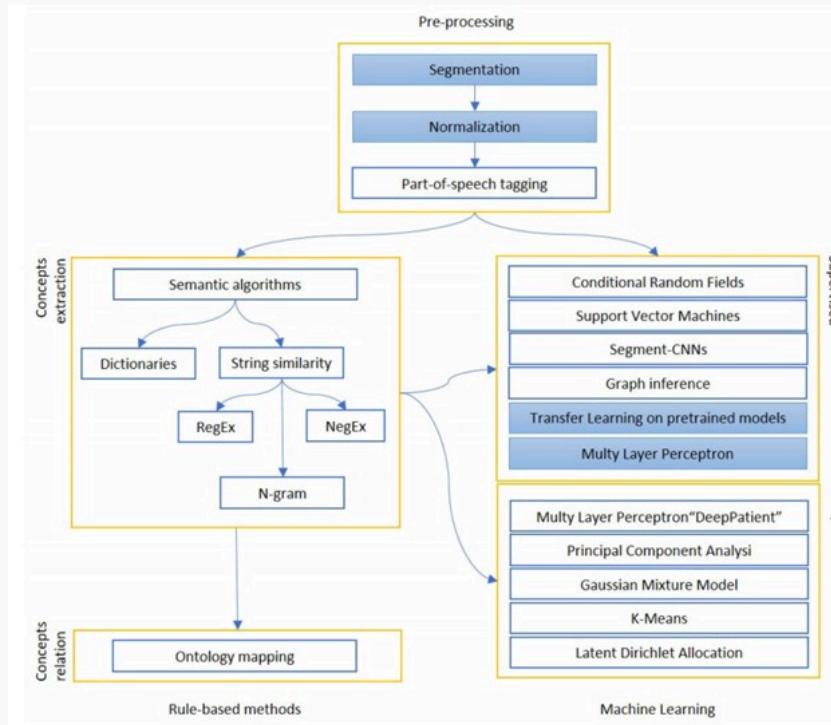
Como sabemos, o **PLN tem grande intersecção com a área de Machine Learning (ML)**, conforme Figura 1. Isso ocorre, pois, os modelos de ML são essenciais para diversas tarefas de PLN como o *POS-Tagging* e *Named Entity Recognition*. E por se utilizar de vários métodos de ML, as abordagens de desenvolvimento são bem similares, conforme apresentado na Figura 2.

Figura 1: Áreas de estudo envolvidas na extração de dados do PEP



A figura mostra as técnicas de extração de informações mais relevantes disponíveis na literatura. Sempre há uma primeira etapa de pré-processamento, onde os documentos são convertidos em tokens e classificados de maneira diferente. Uma abordagem é extrair entidades conhecidas com base em padrões, com métodos baseados em regras. Essas entidades podem ser combinadas a uma ontologia existente ou usadas para criar uma, ou pretendidas como uma primeira etapa para um algoritmo de aprendizado de máquina, que pode ser supervisionado (principalmente classificação) ou não supervisionado (agrupamento).

Figura 2: Abordagens de desenvolvimento na extração de dados do PEP



Fonte: ASSALE, Michela et al. The Revival of the Notes Field: Leveraging the Unstructured Content in Electronic Health Records. *Frontiers in Medicine*, [s. l.], v. 6, p. 66, 2019.

O grande diferencial em se explorar o PLN em dados clínicos reside justamente na **complexidade e particularidades deste tipo de dado**, características estas que já foram exploradas anteriormente na disciplina e na literatura médica (e.g., falta de estrutura formal, erros ortográficos e abreviaturas frequentes, heterogeneidade, variabilidade, dados faltantes, informações implícitas e/ou contraditórias).

Desenvolvendo tarefas essenciais de PLN para extrair informações de narrativas clínicas

Vamos então, ver na prática como são desenvolvidos modelos para processar e extrair dados de narrativas clínicas de pacientes. No vídeo a seguir iremos trabalhar com os algoritmos que chamamos de *Sequence Labeling*. Realizaremos o **POS-Tagging e Named Entity Recognition em texto clínicos**, ou seja, vamos definir as etiquetas morfológicas das palavras (e.g., substantivo, adjetivo, verbo) e identificar as classificações semânticas dos conceitos médicos em meio ao texto (e.g., doenças, medicamentos, procedimentos, sintomas).

POS-Tagging e Named Entity Recognition para textos clínicos

Nesta videoaula iremos acompanhar a aplicação das tarefas de POS-Tagging e NER para textos de pacientes.

POS-Tagging e named entity recognition para textos clínicos



Referências Bibliográficas

COLICCHIO, T. K. Introdução à informática em saúde: fundamentos, aplicações e lições aprendidas com a informatização do sistema de saúde americano. Porto Alegre: Artmed, 2020. [Minha Biblioteca].

JURAFSKY, Dan; MARTIN, James H. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. Third edition draft. 2021. Disponível em: <https://web.stanford.edu/~jurafsky/slp3/>. Acesso em: 01 jan. 2021.

ASSALE, Michela et al. The Revival of the Notes Field: Leveraging the Unstructured Content in Electronic Health Records. *Frontiers in Medicine*, [s. l.], v. 6, p. 66, 2019.

JUHN, Young; LIU, Hongfang. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. *Journal of Allergy and Clinical Immunology*, [s. l.], v. 145, n. 2, p. 463–469, 2020.

SPASIC, Irena; NENADIC, Goran. Clinical Text Data in Machine Learning: Systematic Review. *JMIR Medical Informatics*, [s. l.], v. 8, n. 3, p. e17984, 2020.

DALIANIS, Hercules. Characteristics of Patient Records and Clinical Corpora. In: DALIANIS, Hercules. *Clinical Text Mining*. Cham: Springer International Publishing, 2018. p. 21–34. E-book. Disponível em: http://link.springer.com/10.1007/978-3-319-78503-5_4. Acesso em: 01 jan. 2021.



© PUCPR - Todos os direitos reservados.