



IA Aplicada à Saúde

UNIDADE 02

Caracterização e complexidade dos dados em saúde

| Caracterização e complexidade dos dados em saúde

Para entendermos as particularidades dos dados em saúde, bem como, compreendermos as atividades que dependem da utilização destes dados, vamos efetuar a leitura do capítulo de livro descrito a seguir, para após isso discutirmos e complementarmos nossas descobertas.

Para se aprofundar mais no assunto, leia o capítulo 2 - A Complexidade dos dados em Saúde - do livro Introdução à informática em saúde: fundamentos, aplicações e lições aprendidas com a informatização do sistema de saúde americano.

No capítulo do livro mencionado, muito se fala dos dados que são armazenados no Prontuário Eletrônico do Paciente (PEP), que de fato são os mais importantes e relevantes quando falamos de assistência ao paciente. Porém, há ainda um outro subconjunto de dados que podemos utilizar no contexto da saúde, conforme exposto na seção a seguir.

Dados clínicos vs. biomédicos

Para fins de consistência e normalização de nosso material, vamos aqui estabelecer algumas definições acerca dos dados. Quando falarmos de **dados de saúde** estamos nos referindo a todo o conjunto de dados relacionados ao tema da disciplina. Estes dados são divididos em alguns subconjuntos, que definiremos a seguir.

Os **dados clínicos** serão considerados todos dados associados a um indivíduo (i.e., paciente), ou seja, toda informação acerca do paciente que é de alguma maneira armazenada em seu prontuário, seja este dado estruturado ou não-estruturado.

Já os **dados biomédicos**, representam todo conhecimento relacionado à saúde disponível na literatura, que estão dispostos em sua maioria em forma de artigos científicos e dados coletivos divulgados por organizações e instituições governamentais.

É importante que tracemos esta linha entre estes subconjuntos de dados da saúde, pois o processamento dos mesmos se diferem muito em complexidade e aplicabilidade, situação que será demonstrada no decorrer da disciplina.

Dados estruturados vs. não-estruturados

Independentemente dos subconjuntos apresentados anteriormente, os dados em saúde ainda podem ser classificados como estruturados e não-estruturados. Sendo os **estruturados** aqueles que representam variáveis categóricas (e.g., gênero, raça, plano de saúde) ou numéricas, que são armazenadas de maneira tabular. Já os dados **não-estruturados** incluem os dados de imagens, sinais, voz e texto, que exigem algum tipo de processamento para estruturação, organização e interpretação dos mesmos.

O dado **estruturado** facilita seu processamento por quaisquer algoritmos, inclusive os de IA, podendo ser útil em todos âmbitos da saúde, desde a tomada de decisão, passando por questões administrativo-financeiras e também pela área de pesquisa (e.g., estudos epidemiológicos).

Quando falamos de dado **não-estruturado**, sempre temos um fator extra de complexidade na recuperação das informações. Por exemplo, os exames de imagens devem ser processados por algum algoritmo para extrair dali informações acerca do diagnóstico e condição do paciente.

Alguns dados podem estar em formato estruturado ou não-estruturado, dependendo do sistema de PEP que é utilizado, como os medicamentos prescritos ao paciente, por exemplo. Há PEPs que permitem inclusão destes dados de maneira estruturada (campos separados para nome do medicamento, dose, via de administração e frequência de uso), já outros em que o médico faz a prescrição diretamente em suas narrativas clínicas, em forma textual.

São diversos os dados armazenados no prontuário do paciente, independentemente de sua estrutura. A seguir uma listagem que cobre a maior parte dos dados.

Dados demográficos



EXEMPLO

Idade, Gênero, Plano de saúde, Idioma, Religião, Estrutura familiar, Endereço, Emprego, etc.

Medidas



EXEMPLO

Altura, Peso, Pulso, Frequência respiratória, Temperatura, etc.

Medicamentos



EXEMPLO

Prescrições, Medicamentos sem prescrição, Drogas ilegais, álcool, etc.

Exames laboratoriais



EXEMPLO

Componentes sanguíneos, Urina, Fezes, Saliva, Líquido espinhal, Pulmões, etc.

Patologia



EXEMPLO

Examinação qualitativa e quantitativa de qualquer tecido corporal (e.g., biópsia de amostras).

Medições a nível celular.

Microbiologia



EXEMPLO

Teste de sensibilidade a antibióticos

Narrativas clínicas



EXEMPLO

Sumário de alta

Notas ambulatoriais

Evolução de enfermagem

Notas de especialista (e.g., laudos radiológicos, notas de nutrição)

Carta de encaminhamento.

Cobrança



EXEMPLO

Diagnósticos (geralmente codificados com a CID – Classificação Internacional de Doenças)

Procedimentos

Administração



EXEMPLO

Serviços prestados

Transferências



EXEMPLO

Raio-X

Ultrasound

Tomografia

PET-scan

Magnetic resonance imaging

Endoscopy

Photographs

Vital signs and monitoring



EXEMPLO

Steps

Change in elevation

Physical exercises



EXEMPLO

Frequência cardíaca

Frequência respiratória

Temperatura

Pressão

Peso

Dieta

Glicemia

Alergias

Atenção

Humor

Sono

Dor

Narrativas Clínicas

Dentre os dados mais importantes do PEP estão as narrativas clínicas, que são **dados textuais, geralmente escritos como texto-livre, sem uma padronização ou formato estabelecido**. Nelas, o corpo clínico (e.g., médicos, enfermeiros, psicólogos, fisioterapeutas) descreve informações acerca do histórico de saúde do paciente como diagnósticos, procedimentos realizados, sintomas, continuidade do cuidado, entre outras informações essenciais ao cuidado do paciente, e que podem ter ainda usos secundários, como a pesquisa clínica.

Dentre as narrativas podemos encontrar: **sumários de alta, evoluções de enfermagem, notas ambulatoriais, laudos radiológicos**, entre outros (exemplos serão apresentados nas seções a seguir). A partir da interpretação destes textos, é possível compreender todo *timeline* do paciente, o que é essencial não apenas para a tomada de decisão médica, mas também para a compreensão de uma série de fatores associados ao cuidado ao paciente.

A seguir são apresentados três exemplos de narrativas:

- Um **sumário de alta** de um paciente com suspeita de doença coronariana e diversas outras comorbidades, escrito pelo médico no momento da alta do paciente, após sua internação no hospital.
- Uma **nota ambulatorial** de atendimento de um paciente com doença renal crônica e outras comorbidades associadas, escrita pelo médico no momento do atendimento no ambulatório.
- Uma **evolução de enfermagem**, escrita pela equipe de enfermagem para descrever o status de monitoração do paciente enquanto internado no hospital.



EXEMPLO

Sumário de alta de pacientes

PACIENTE DIABÉTICA, COM IMC ELEVADO, HAS, TABAGISTA PESADA ATIVA, USUÁRIA CRONICA DE CORTICÓIDE 920MG /DIA PREDNISONA) POR A.REUMATÓIDE, CAR 2001 COM LESÕES 3 VASOS MODERADAS (<30%), INTERNA POR DOR TORACICA ATÍPICA EM REPOUSO, TRATADA COMO SINDROME CORONARIANA AGUDA PELAS VARIAVEIS CLÍNICAS (DOENÇA CORONARIANA PRÉVIA DEMONSTRADA) + FATORES DE RISCO MÚLTIPLOS PARA ATROSCLEROSE. SEM ALT NOVAS ECG DA CHEGADA E SEM CURVA ENZIMATICA TÍPICA DE IAM. FOI OPTADO POR MANEJO CONSERVADOR, SENDO QUE A PACIENTE APRESENTOU MELHORA IMPORTANTE DA DOR COM TRATAMENTO CLÍNICO OTIMIZADO, TEM ALTA EM BOM ESTADO GERAL, ASSINTOMÁTICO, COM USO DE TODAS AS INTERVENÇÕES FARMACOLÓGICAS COM BENEFÍCIO COMPROVABDO POR ENSAIOS CLÍNICOS RANDOMIZADOS E COM RETORNO AMBULATORIAL JÁ MARCADO PARA MED INTERNA ONDE JÁ REALIZAVA ACOMPANHAMENTO. SUGERIMOS PARA MÉDICO ASSITENTE REDUZIR DOSES CORTICÓIDES E POSSÍVEL.
(Autor, 2021)

O **sumário de alta** é um documento que descreve os dados mais relevantes acerca da internação do paciente, sendo essencial para continuidade do cuidado. Nele há um histórico completo acerca da condição do paciente, procedimentos realizados, medicamentos prescritos, entre outros. Este documento pode inclusive servir de apoio às equipes médicas de outras instituições, para dar continuidade ao atendimento.



EXEMPLO

Nota ambulatorial de atendimento ao paciente

Paciente com 70 anos, branca, viúva, 2 filhos.

HMA:

- Consulta de retorno ao ambulatório de nefro - Diabética, HAS, DRC estagio 4 - há 8 anos - não dialítica.

HMP:

- DM 2 - há 8 anos

- HAS - há 8 anos

- Hipotireoidismo - há 8 anos

- Insuficiencia cardíaca?

- Perda de acuidade visual, retinopatia diabética (fotocoagulação)

- Medicação - Furosemida 40 mg - 2 cp manha e 2 apos almoço, levotoroxina 100 mcg - 2 cp manha, insulina NPH 20 UI noite, enalapril 10 mg 2 cp manha e noite, AAS 100 mg - 1cp - almoço, simvastatina 20 mg 2 cp noite, atenol 50 mg - 1cp manha e noite, espironolactona 25 mg 1 cp manha, clonidina 0,1 mg 1 cp manha e noite.

- Glicemias noturnas alteradas (alta), hipoglicemia de manhã esporadicamente.

- CX anteriores: Colecistectomia há 10 anos

- Crise hipertensiva (AVCI?) há 10 anos

- Operação para catarata há 4 meses

- Nega alergia à medicamentos

*Clearence: 25

- Creatinina: 2

- Glicose: 113

- Ureia: 117

- Calcio total: 7,7

- HDL: 37

- Fósforo: 5,5

- PTH: 294,1

- Hb: 11,4

- Leucocitos: 10310

- No parcial de urina: proteinas 1000, leucocitos 42

HMF: Hx familiar de HAS,DM , nega neoplasia na familia.

CHV: Alimenta-se bem, alimentação variada, dorme bem. Nega tabagismo e etilismo.

ROS: Nega queixas em demais órgãos e aparelhos, além dos citados acima.(Autor, 2021)

A **nota ambulatorial** é o documento que contém todas informações relevantes de um paciente quando o mesmo realiza uma consulta médica no ambulatório/clínica, ou seja, sem internação hospitalar. Nele podem conter não apenas as queixas do paciente, mas também resultados de exames físicos e laboratoriais, histórico familiar, diagnósticos, tratamentos, entre outras informações importantes sobre a condição do paciente.



EXEMPLO

Evolução de enfermagem

7:45Hrs: Cliente em Ramsay 6, pupilas iso / foto +, intubado em TOT nº 8.0 rima labial 24cm DIR, VM nos parâmetros CMV, VC 540ml, peep 5cmH₂O, FR 20, FiO₂ 40%, face corada, mucosa oral hidratada, boa perfusão periférica, no leito com cabeceira elevada a 45°. Jejum até 2º ordem. Mantém com monitorização de multiparâmetros: ECG, PAM em radial ESQ (oxilante), SpO₂, SNG aberta em narina ESQ com estase gástrica ++ de aspecto biliosa, catéter venoso central duplo-lúmen em subclávia ESQ com noradrenalina 14ml/hr, dormonid 35ml/hr (planos dobrados) em curso BI, acesso venoso periférico em MSE com fentanil 20ml/hr (plano dobrado) em curso BI, DTF D anterior com débito seroso e posterior sero-hemático, DTF E com débito seroso, ctvos em regiões toracoabdominal + dorsal com presença de secreção serosa em pouca quantidade. Apresenta ausculta cardíaca rítmica e normofonética, ausculta pulmonar MV + sem R/A, abdômen plano e flácido com RHA hipoativos, edemas em MMSS cacofo I. Faz uso de SVD aberta com débito diurético presente + de coloração amarelo claro, eliminação fecal ausente. Seguem cuidados intensivos de enfermagem.(Autor, 2021)

Por fim, a **evolução de enfermagem**, que ao contrário das outras narrativas clínicas, é escrita pela equipe de enfermagem, durante a internação do paciente. Neste documento são descritos todos cuidados ao paciente realizados pela equipe de enfermagem, bem como monitoramento de todos sinais importantes à condição do paciente.

RELEMBRANDO: Anteriormente você foi apresentado à vários casos de uso na aplicação de IA para saúde, ligadas a **melhoria do cuidado ao paciente**, através de sistemas de **apoio a decisão, otimização de processos médico-administrativos**, e até mesmo **apoio à pesquisa e indústria farmacêutica**. Dentre estas oportunidades, diversas estão atreladas ao processamento das narrativas clínicas, que muitas vezes são o único local onde determinadas informações são armazenadas.



REFLEXÃO

Mas se os dados das narrativas clínicas são tão importantes e podem apoiar tantas iniciativas, por que eles ainda são pouco explorados, principalmente no cenário brasileiro?

Isso se deve a uma série de fatores, como a complexidade de extrair e interpretar informações de dados textuais, falta de recursos semântico-linguísticos para o idioma português e questões de ética e privacidade no acesso à dados de pacientes.

Desafios e dificuldades no processamento de textos clínicos

A **falta de estrutura formal e escrita livre** são fatores dificultadores quando pensamos em extrair dados de narrativas clínicas, pois a maneira como os dados podem ser escritos e formatados nas narrativas pode variar muito de profissional para profissional. A **heterogeneidade e variabilidade** dos textos também influem na complexidade de manipulação dos dados, pois os documentos podem variar muito em tamanho (curto vs. comprido), completude (informações completas vs. faltantes), fase do cuidado ao paciente (primeira visita, acompanhamento, alta) e até mesmo o estágio da doença (e.g., doença renal crônica grau 1 vs. paciente transplantado vs. paciente terminal), que pode acarretar em diferentes conjuntos de informações relevantes.

A alta frequência de uso de **abreviaturas** e **acrônimos** não oficiais, bem como a presença de **erros ortográficos** também impactam no processamento dos dados, podendo ocasionar erros de extração e interpretação das informações.

Outra característica muito presente em narrativas clínicas são as **negações** e uso de **linguagem especulativa**, o que pode alterar completamente o sentido da informação extraída. Por exemplo, se estivéssemos procurando por pacientes com diabetes em nosso sistema de PEP, caso utilizássemos apenas uma busca simples pela palavra “diabetes”, textos que contém a frase “paciente nega diabetes” também seriam selecionados. Muitas vezes o profissional também escreve um diagnóstico em forma de hipótese, para documentar a investigação de um caso clínico, estes casos podem ocasionar o mesmo problema citado acima.

Um desafio adicional é a captura de **informações implícitas** no texto clínico, que podem ser inferidas através do uso de outros dados. Por exemplo, podemos saber que o paciente tem condição X, sem ela estar descrita, apenas utilizando alguns sintomas e resultados de exames relacionados no prontuário. Além disso, devemos levar em conta algumas **informações contraditórias** no texto, ocasionadas muitas vezes pelo hábito do copiar-e-colar da equipe médica no momento de construir sua narrativa.

IMPORTANTE

Você pode encontrar elementos de texto que comprovam a presença de alguns dos aspectos textuais apresentados acima (uso de



abreviaturas, negações, erros ortográficos, variabilidade, falta de estrutura formal), nos exemplos de narrativas apresentados na seção anterior.

Como pode-se ver, as características das narrativas clínicas já são um grande desafio para algoritmos de **Processamento de Linguagem Natural** (PLN), que em muitos casos necessitam de técnicas robustas de **Machine Learning** para superar toda variabilidade textual e ter um modelo generalizável para solução do problema. Modelos estes que em sua maioria se apoiam em Aprendizado Supervisionado, ou seja, necessitam de dados rotulados para seu treinamento. E uma das grandes dificuldades quando trabalhamos com IA em Saúde é justamente na **obtenção dos dados de pacientes**, devido a todas questões regulatórias de privacidade e segurança da informação (discutiremos isto mais a fundo no decorrer da disciplina). Ou seja, temos um cenário em que a **falta de dados disponíveis de maneira aberta para a comunidade científica** dificultam a construção de soluções ótimas para extração de dados em narrativas clínicas, impactando não apenas o treinamento de novos modelos, mas também na impossibilidade da realização de um *benchmarking* (i.e., comparação dos resultados) entre diferentes soluções.

A maioria absoluta dos recursos semântico-linguísticos associados ao processamento de textos clínicos estão no idioma inglês, porém, recentemente algumas iniciativas iniciaram a disponibilização de modelos e corpora para o idioma português. Destaca-se o grupo **HAILab** (*Health Artificial Intelligence Lab*) da PUCPR, que compartilhou uma série de modelos para manipulação de textos clínicos.



SAIBA MAIS

Quer saber mais sobre o HAILab, [clique aqui](#).

Padronização dos dados e terminologias em saúde

A **padronização dos dados** contidos no PEP pode auxiliar na utilização dos dados em saúde e permitir que todas **informações sejam corretamente interpretadas** por todos envolvidos no processo de cuidado, suportando a **reusabilidade** dos dados, melhorando a **eficiência** dos serviços de saúde e **evitando erros** ao reduzir esforços duplicados na entrada dos dados.

Para termos dados padronizados precisamos definir: (i) o conteúdo a ser armazenado, (ii) as terminologias utilizadas para representar os dados, (iii) como os dados serão intercambiados, e (iv) como o conhecimento (e.g., diretrizes e protocolos clínicos) será representado no Sistema de Informação em Saúde.

Em nossa disciplina iremos focar mais nas terminologias, como a **UMLS**, **CID-10** e **SNOMED CT**, que serão utilizadas em nossas atividades práticas. Para saber mais sobre as questões de interoperabilidade de dados em saúde você pode consultar o “Capítulo 6: Padrões de comunicação em Saúde”, de nossa bibliografia principal.

Para se aprofundar mais no assunto leia os trechos da página 52 até 54 e da página 60 até 63 do Capítulo 6- Padrões De Comunicação Em Saúde - do livro Introdução à informática em saúde: fundamentos, aplicações e lições aprendidas com a informatização do sistema de saúde americano.

Como funcionam as terminologias médicas utilizadas na padronização dos dados em saúde

Nesta videoaula vamos entender como funcionam as principais terminologias médicas, como elas são utilizadas em Sistemas de Informação em Saúde e como elas auxiliam os algoritmos de IA.

Terminologias médicas



A partir dos aprendizados desta Unidade, nós:

- Entendemos as particularidades dos dados em saúde;
- Diferenciamos dados clínicos de biomédicos;
- Definimos os dados estruturados e não-estruturados;
- Listamos os dados disponíveis no Prontuário Eletrônico do Paciente;
- Detalhamos a importância, características e complexidade no processamento de narrativas clínicas;
- E finalmente, fomos introduzidos às terminologias médicas utilizadas para padronizar os dados em saúde.

| 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].

SHORTLIFFE, E. H., BARNETT, G. O. Biomedical Data: Their Acquisition, Storage, and Use. Springer, New York, NY, 2006.



© PUCPR - Todos os direitos reservados.