

به نام خدا



موضوع:
پردازش زبان طبیعی (NLP)

نام درس: هوش مصنوعی

نام ارائه دهنده: مهسا یزدان فر

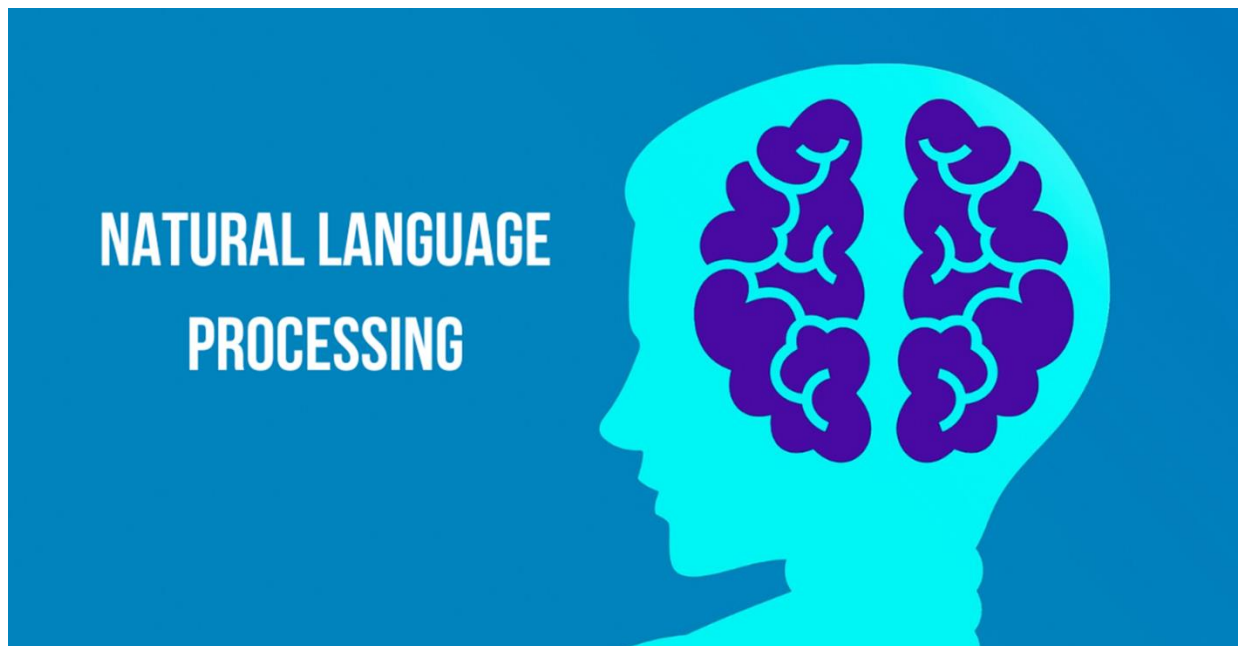
تابستان 1403

عنوان ها:

- پردازش زبان طبیعی چیست؟
- اهمیت نیاز به پردازش زبان طبیعی
- آزمون تورینگ
- تکنیک های پردازش متن
- کاربرد های NLP در حوزه متن
- نتیجه گیری
- منابع

پردازش زبان طبیعی (NLP) چیست؟

تا به حال شده به این فکر کنید که موتور جستجوگر گوگل چگونه کار میکند؟ چطور غلط های املایی تصحیح می شوند؟ یا شده فکر کنید که دستیاران صوتی گوگل، سیری و الکسا چگونه صحبت مارا به متن تبدیل می کنند و سریعاً به ما پاسخ می دهند؟ یا مثلاً بگویید کاش برنامه ای وجود داشت که مقال طولانی را سریعاً به صورت خلاصه، مختصر و مفید برایمان توضیح می داد؟ تمام این کار ها با NLP امکان پذیر است پس بیایید NLP را بیشتر بشناسیم.



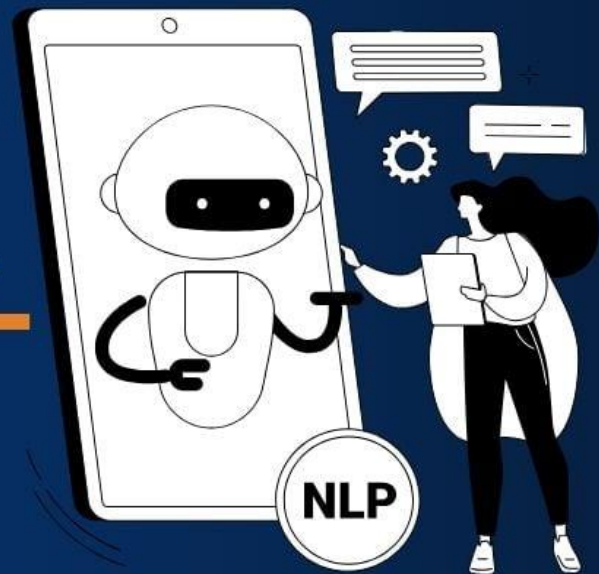
NLP چیست؟

پردازش زبان طبیعی یا (NLP (Natural Language Processing شاخه‌ای از هوش مصنوعی است که به ماشین‌ها این امکان را می‌دهد تا زبان‌های رایج میان انسانها را بخوانند، داده‌ها را درک کند و سپس از آنها معنی استخراج کنند. NLP به تعامل بین کامپیوتر و زبان‌های طبیعی یا انسانی می‌پردازد. بنابراین NLP بر ارتباط میان انسان و رایانه متمرکز است.

چالش اصلی در این زمینه، ماشینی کردن فرایند درک و برداشت مفاهیم بیان شده با یک زبان طبیعی انسانی است. به تعریف دقیق تر عبارت است از استفاده از کامپیوتر برای پردازش زبان گفتاری و نوشتاری. به این معنی که رایانه‌ها را قادر سازیم تا گفتاریا نوشتار تولید شده در قالب و ساختار یک زبان طبیعی را تحلیل و درک نموده و یا اینکه آن را تولید کنند. در این صورت میتوان به ترجمه زبان ها پرداخت.

هدف اصلی در NLP، ایجاد تئوری‌های کحاسباتی از زبان، با استفاده از الگوریتم‌ها و ساختارهای داده‌ای موجود در علوم رایانه است. روش کلی کار به این صورت است که NLP زبان شناسی و مدل سازی مبتنی بر قوانین زبان انسانی را با مدل‌های آماری، machine learning و deep learning ترکیب می‌کند. این فناوری‌ها با هم، رایانه‌ها را قادر می‌سازند تا زبان انسان را در قالب متن پردازش کنند و در نهایت معنای دقیق را درک کند. NLP زمینه‌ای است که بر تعامل بین علم داده و زبان طبیعی میان انسان‌ها تمرکز می‌کند.

What is Natural Language Processing?



اهمیت نیاز به پردازش زبان طبیعی

هر چیزی که ما انسانها بیان می‌کنیم (چه به صورت شفاهی و چه به صورت نوشتاری) حاوی حجم عظیمی از اطلاعات است. موضوعی که انتخاب می‌کنیم، لحن ما، کلمات ما و... همگی شکلی از داده است که می‌تواند تفسیر شود و از آن داده‌ها را استخراج کرد که ما می‌توانیم با استفاده از این اطلاعات رفتارها و اطلاعات را درک کرده و حتی پیش‌بینی کنیم. اما یک مشکل وجود دارد: کامپیوترها اعداد و داده‌های جدولی را به خوبی درک می‌کنند اما یک نفر ممکن است صدها یا هزاران کلمه را ایجاد کند و یا هر جمله را با پیچیدگی مخصوص به خود بسازد. و زبانی که ما روزانه از آن استفاده می‌کنیم و صحبت می‌کنیم دارای ساختار مشخصی نیست مانند داده‌های تولید شده از مکالمات روزمره، توییت‌ها و... این داده‌ها، داده‌های بدون ساختار هستند و اغلب داده‌های موجود در دنیای واقعی در این دسته قرار می‌گیرند ولی ما انتظار داریم کامپیوتر این جملات و داده‌های بدون ساختار را به خوبی درک کند و اطلاعات را از آن‌ها استخراج کند. برای پردازش این نوع از داده به NLP نیاز داریم.

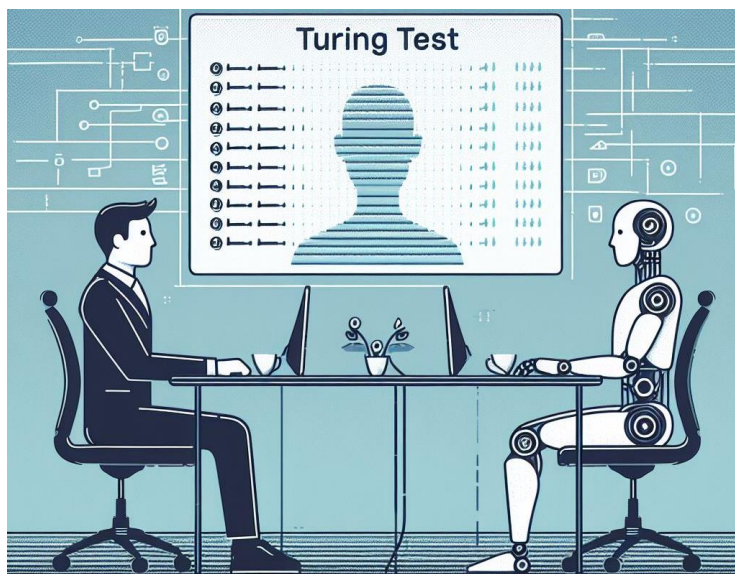
این کار انقدر بزرگ و مهم است که خیلی‌ها اعتقاد دارند آخرین گام هوش مصنوعی این است که سیستم‌ها با ما حرف بزنند و ما نتوانیم تشخیص دهیم که به جای یک انسان با کامپیوتر ارتباط برقرار کرده ایم. این کار توسط آزمون تورینگ انجام می‌شود.



آزمون تورینگ

آزمون تورینگ، آخرین مرحله‌ی تست هوشمندی ماشین در زمینه‌ی **NLP** است. این آزمون به این شکل است که یک شخص شروع به چت کردن با یک ماشین می‌کند در حالی که نمی‌داند مخاطب او یک انسان نیست. اگر او نتواند در پایان صحبتش تشخیص دهد که با یک ماشین چت کرده، ماشین در این آزمون موفق شده است.

این آزمون امروزه به عنوان ملاک هوشمندی شناخته می‌شود و این معیار همانطور که گفتیم به توانایی یک ماشین برای جعل هویت انسان در یک مکالمه مطلوب در لحظه می‌پردازد.



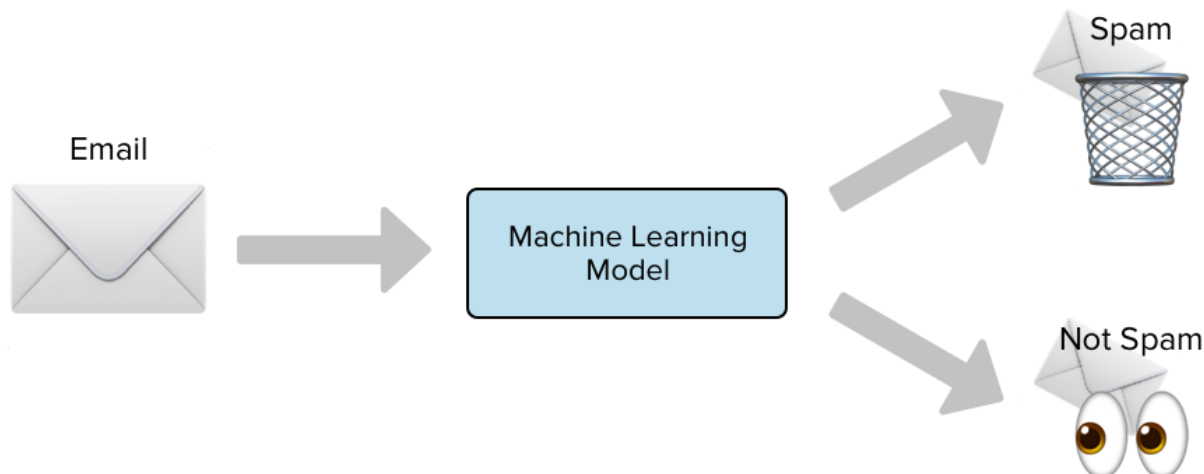
کاربردهای NLP در حوزه متن

NLP به برنامه‌های کامپیوتری کمک می‌کند تا متن را از یک زبان به زبان دیگر ترجمه کنند، به مکالمات متنی پاسخ دهند و حجم زیادی از متن را به سرعت خلاصه کنند (حتی به صورت در لحظه یا real time). دستیارهای دیجیتال، نرم‌افزار تصحیح، چت‌ربات‌های خدمات مشتری و سایر امکانات رفاهی مبتنی بر متن همگی از NLP استفاده می‌کنند.



تشخیص spam

در نگاه اول ممکن است تشخیص اسپم یا هرزنامه را به عنوان یکی از کاربردهای NLP در نظر نگیرید، اما بهترین فناوری‌های حال حاضر (برای مثال گوگل در بخش جیمیل) برای تشخیص هرزنامه از قابلیت‌های طبقه بندی متن با کمک NLP استفاده می‌کنند. برخی از شاخص‌های طبقه بندی متن عبارتند از استفاده بیش از حد از برخی از اصطلاحات تبلیغاتی، گرامر بد، زبان تهدیدآمیز، موضوع نامناسب، نام شرکت‌ها با املای اشتباه و...



ترجمه ماشینی

گوگل ترنسلیت (Google Translate) نمونه‌ای از کاربرد مستقیم NLP است که به طور گسترده در دسترس همه قرار دارد. ترجمه ماشینی چیزی بیشتر از جایگزینی کلمات یک زبان با کلمات زبانی دیگر است. یک ترجمه مناسب باید معنی و لحن زبان ورودی را به دقت دریافت کرده و آن را به متنی با همان معنا و تاثیر دلخواه در زبان دوم ترجمه کند. ابزارهای ترجمه ماشینی از نظر دقت پیشرفت خوبی دارند. یک راه عالی برای آزمایش هر ابزار ترجمه ماشینی، ترجمه متن به یک زبان و سپس ترجمه مجدد خروجی به زبان اصلی است.



چت بات‌های گفتگو

دستیارهای مجازی مانند سیری در سیستم عامل اپل و الکسا در آمازون از تشخیص گفتار برای تشخیص الگوهای دستورات صوتی و تولید زبان طبیعی استفاده می‌کنند تا با اقدامات مناسب یا نظرات مفید پاسخ دهند. چت بات‌ها یاد می‌گیرند تا سرنخ‌های متنی را در درخواست‌های انسانها تشخیص دهند و از آنها برای ارائه پاسخها یا گزینه‌های بهتر در طول زمان استفاده کنند.



خلاصه سازی متن

خلاصه سازی متن یکی از تکنیک های NLP برای هضم حجم عظیمی از متن دیجیتالی و ایجاد خلاصه هایی برای پایگاه های اطلاعاتی تحقیقاتی یا خوانندگان پر مشغله که وقت کافی ندارند استفاده میکند. بهترین برنامه های خلاصه سازی متن از استدلال معنایی و تولید زبان طبیعی برای تولید متن با توجه به زمینه متن و نتیجه گیری و جمع بندی استفاده می کند.



ابهام زدایی

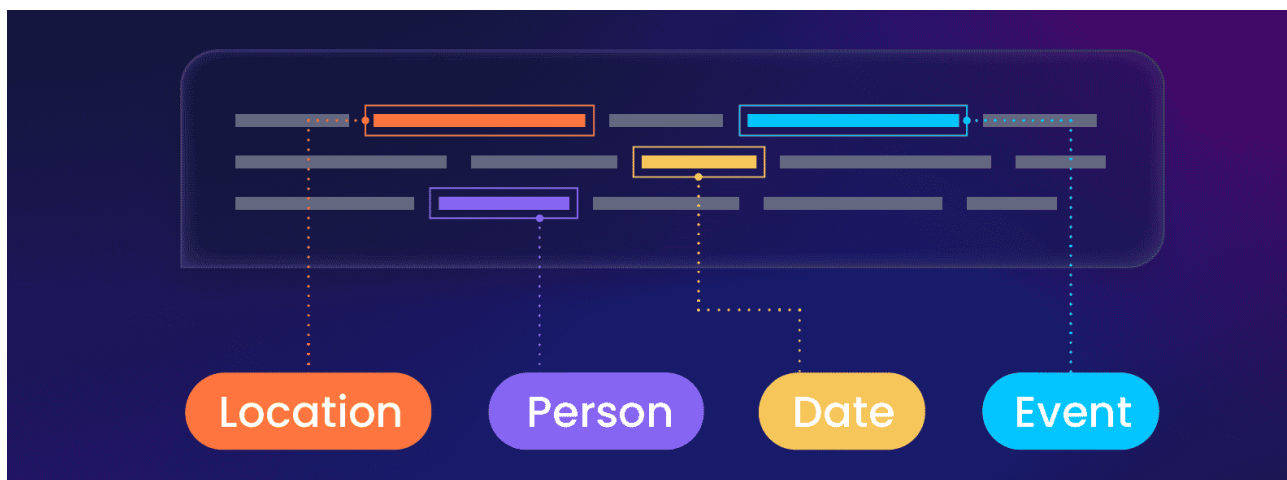
فرایند ابهام زدایی عبارت است از انتخاب معنای یک کلمه از میان معنای چندگانه از طریق رویکرد تحلیل معنایی. برای این اساس کلمه ای انتخاب می شود که معنایش بیشترین انطباق بر اساس متن دارد. برای مثال ابهام زدایی از معنای کلمات موجود در متن استفاده می کند تا معنای کلمه شیر را تشخیص دهد.

Disambiguation

Why It's Important for NLP

شناسایی موجودیت

کلمات یا عبارات را به عنوان موجودیت های مفید شناسایی می کند. برای مثال شیراز را به عنوان یک مکان و علی را به عنوان یک مرد شناسایی می کند.



NLP در پزشکی

NLP تشخیص و پیش بینی بیماری ها را بر اساس پرونده الکترونیکی سلامت و گفتار خود بیمار امکان پذیر می کند. به عنوان مثال Comprehend medical یکی از سرویس های آمازون از NLP برای استخراج شرایط بیماری داروها و نتایج درمان و سایر سوابق سلامت الکترونیک استفاده می کند.



نتیجه

NLP یکی از حوزه های در حال رشد و امیدوار کننده در هوش مصنوعی است. در حال حاضر این فناوری در بسیاری از برنامه های کاربردی که ما روزانه از آن ها استفاده می کنیم همانطور که در این ارائه اشاره شد حضور دارد. از چت بات ها گرفته تا موتور های جست و جو به لطف NLP مشاغل برخی از فرایندهای روزانه ی خود را به صورت خودکار انجام می دهند و از داده های بدون ساختار خود حداکثر استفاده را می کنند.



<https://www.hamyarit.com/blog/nlp/>

<https://7learn.com/blog/what-is-nlp>

https://fa.wikipedia.org/wiki/%D9%BE%D8%B1%D8%AF%D8%A7%D8%B2%D8%B4_%D8%B2%D8%A8%D8%A7%D9%86%E2%80%8C%D9%87%D8%A7%DB%8C_%D8%B7%D8%A8%DB%8C%D8%B9%DB%8C

<https://magfa.com/nlp/>

<https://rahaco.net/mag/nlp/>