

கடந்த சில ஆண்டுகளாக நியூரோ லிங்விஸ்டிக் ப்ரொக்ரம் இன் பரிணாமம்(NLP)

ஒரு கணினி நிரலின் இயல்பான மொழியைப் புரிந்துகொள்ளும் திறன் அல்லது மனித மொழியைப் பேசும் மற்றும் எழுதும் திறன், இயற்கை மொழி செயலாக்கம் (NLP) என அழைக்கப்படுகிறது. இது இயந்திர நுண்ணறிவின் (AI) ஒரு பகுதியாகும். NLP 50 ஆண்டுகளுக்கும் மேலாக உள்ளது மற்றும் மொழியியல் வேர்களைக் கொண்டுள்ளது. பெருநிறுவன நுண்ணறிவு, தேடுபொறிகள் மற்றும் மருத்துவ ஆராய்ச்சி உள்ளிட்ட பல்வேறு தொழில்களில் இது பல நடைமுறை பயன்பாடுகளைக் கொண்டுள்ளது.



இயற்கை மொழி செயலாக்கத்தின் செயல்முறை என்ன?

மனித மொழி இயற்கையான மொழி செயலாக்கத்தில் துண்டுகளாக உடைக்கப்படுகிறது, இதனால் வாக்கிய அமைப்பு மற்றும் வார்த்தையின் அர்த்தத்தை ஆராய்ந்து புரிந்து கொள்ள முடியும். இது

மக்கள் பேசும் அல்லது எழுதப்பட்ட உரையைப் படிக்கவும் புரிந்துகொள்ளவும் கணினிகளுக்கு உதவுகிறது. NLP தொழில்நுட்பங்கள் மனித மொழியைப் புரிந்துகொள்வதற்கு முன், தரவு விஞ்ஞானிகள் முதலில் சில முக்கிய NLP முன் செயலாக்க பணிகளை மேற்கொள்ள வேண்டும்:

- டோக்கனைசேஷன்: உரை சிறிய சொற்பொருள் கூறுகளாக அல்லது ஒற்றை உட்பிரிவுகளாகப் பிரிக்கப்படும்போது டோக்கனைஸ் செய்யப்படுகிறது.
- பேச்சின் ஒரு பகுதி-குறியிடல்: பெயர்ச்சொற்கள், வினைச்சொற்கள், உரிச்சொற்கள், வினையுரிச்சொற்கள், பிரதிபெயர்கள், முதலியன போன்ற சொற்களைக் குறிக்கவும்.
- லெமடிசேஷன் மற்றும் ஸ்டெம்மிங்: சொற்களஞ்சியத்தை தரப்படுத்த சொற்களை அவற்றின் அடிப்படை வடிவங்களுக்கு குறைத்தல்

வார்த்தை நீக்கத்தை நிறுத்து: முன்மொழிவுகள் மற்றும் அதிக அசல் உள்ளடக்கத்தை பங்களிக்காத கட்டுரைகள் (at, to, a, the) போன்ற பொதுவான சொற்களை நீக்குதல்.

NLP கருவிகள் உரையை ஒரு இயந்திரம் புரிந்துகொள்ளக்கூடிய வடிவமாக மாற்றும்.

NLP இன் நிஜ-உலகப் பயன்பாடுகள்:

மின்னஞ்சல் வடிப்பான்கள்:

கடந்த பத்து ஆண்டுகளில் நீங்கள் மின்னஞ்சல் கணக்கைப் பயன்படுத்தியிருந்தால், நீங்கள் சந்தேகத்திற்கு இடமின்றி NLP தொழில்நுட்பத்தில் லாபம் பெற்றிருக்கிறீர்கள். மின்னஞ்சல் உரைத் தரவுகளில் ஒரு அல்காரிதம் சரியாகப் பயிற்றுவிக்கப்பட்டவுடன், மின்னஞ்சல்களை வழக்கமான மின்னஞ்சல்கள், ஸ்பேம் மின்னஞ்சல்கள் அல்லது தீங்கு விளைவிக்கும் மின்னஞ்சல்கள் என சரியாக அடையாளம் கண்டு, வகைப்படுத்தி, லேபிளிடலாம். இதேபோல், மின்னஞ்சல் வழங்குநரைப் பொறுத்து, வெவ்வேறு மின்னஞ்சல் வடிப்பான் வகைகள், இது போன்ற சமூக அல்லது விளம்பரம், பயன்படுத்தப்படலாம். வலுவான மின்னஞ்சல் வடிகட்டுதல், முக்கியமான தரவுகளின் வெளிப்பாட்டைக் குறைப்பதன் மூலம், யாராவது ஒரு தீங்கிழைக்கும் மின்னஞ்சலைக் கிளிக் செய்து திறக்கும் வாய்ப்பை வியத்தகு முறையில் குறைக்கலாம்.



ஸ்மார்ட் உதவியாளர்கள்:

சிரி, அலெக்சா மற்றும் கோர்டானா போன்ற ஸ்மார்ட் அசிஸ்டென்ட்கள் என்எல்பியின் மிகவும் நன்கு அறியப்பட்ட எடுத்துக்காட்டுகள் எனப் புகழ் பெற்றிருக்கலாம். அவர்கள் NLP ஐப் பயன்படுத்தி வார்த்தையின் தண்டுகள், பேச்சின் பகுதிகள் மற்றும் பிற மொழியியல் அம்சங்களில் மொழியைப் பிரிக்கிறார்கள். மீதமுள்ளவை இயற்கை மொழி உருவாக்கத்தால் (NLG) கையாளப்படும், இது இயந்திரங்களுக்கு "பேசும்" திறனையும், இயற்கையான மொழி புரிதலையும் (NLU) வழங்குகிறது, இது இயந்திரங்களை மொழியைப் புரிந்துகொள்ள உதவுகிறது. வெறுமனே, இது நீங்கள் விரும்பும் எதிர்வினையில் விளைகிறது. அதே கொள்கைகள் NLP-பவர்டு சாட்போட்களுக்கு பொருந்தும், மற்றொரு வகை ஸ்மார்ட் அசிஸ்டென்ட், ஆனால் அவை பேச்சு அங்கீகாரத்திற்கு பதிலாக பயனர்களின் உரை உள்ளீட்டிற்கு பதிலளிக்கின்றன. தகவல்களின் ஆதாரமாக அவற்றின் பயனைக் கருத்தில் கொண்டு, பெரும்பாலான இணைய வணிகங்கள் இப்போது தங்கள் வலைத்தளங்களில் தங்கள் முக்கிய தகவல்தொடர்பு வழிமுறையாக அவற்றைப் பயன்படுத்துகின்றன.



உரை பகுப்பாய்வு:

பல வணிகங்கள் என்ன செய்ய வேண்டும் என்பதை விட அதிகமான தரவுகளைக் கொண்டிருப்பதால், நுண்ணறிவுமிக்க முடிவுகளை எடுப்பது கடினமாக இருக்கும். இதன் காரணமாக, பல நிறுவனங்கள் தங்கள் கட்டமைக்கப்படாத தரவை விளக்குவதற்கு NLP மற்றும் உரை பகுப்பாய்வுகளை நம்பியுள்ளன. பெயர்கள், தேதிகள், நாணய மதிப்புகள் மற்றும் போன் எண்கள் போன்ற முக்கியமான தகவல்களை அடையாளம் காணும் திறனை பயனர்கள் பெற்றுள்ளனர், இதன் மூலம் பெயரிடப்பட்ட நிறுவனம் பிரித்தெடுத்தல் போன்ற அடிப்படை NLP அம்சங்களுக்கு நன்றி. அதன் பிறகு, கோப்புகள், விரிதாள்கள், வலைப்பக்கங்கள் மற்றும் சமூக செய்திகள் உட்பட பல்வேறு அளவுகள் மற்றும் வகைகளின் ஆவணங்களில் இந்த முக்கியமான தகவலைக் கண்டறிவதை எளிதாக்குவதற்கு, நிறுவனங்கள் முன்னமைக்கப்பட்ட வகைகளாகப் பிரிக்கப்படுகின்றன. காப்பீட்டுத் துறையும் கூட கோரிக்கைகள் மற்றும் இடர் மேலாண்மை முடிவெடுப்பதில் உதவ NLP உரை பகுப்பாய்வுகளைப் பயன்படுத்துகிறது.



தேடல் முடிவுகள்:

மக்கள் தற்போது பெரும்பாலும் ஆன்லைன் தேடல்கள் மூலம் தகவல்களை அணுகுகின்றனர். இதன் விளைவாக, அதிகமான வணிகங்கள் தங்கள் மென்பொருளில் NLP தேடல் திறன்களைச் சேர்ப்பதன் நன்மைகளைப் புரிந்துகொள்கின்றன. Expert.ai போன்ற முன்னணி NLP கண்டுபிடிப்பாளர்கள் தேடலை அதன் முழு திறனுக்கும் பயன்படுத்துகின்றனர். வாடிக்கையாளர்கள் தனிப்பயன் நிறுவன தேடல் தீர்வுகளை வடிவமைக்க முடியும், அதாவது செய்தி மூலம் தேடலை இயக்குவது அல்லது ஒரு குறிப்பிட்ட தலைப்புடன்

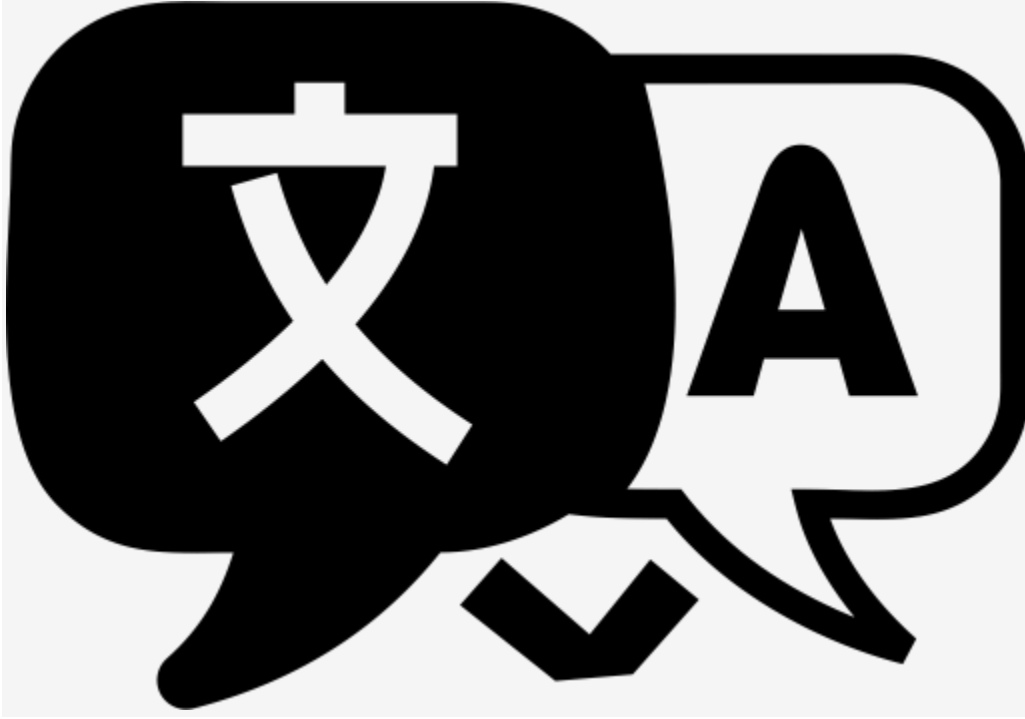
இணைக்கப்பட்ட ஆவணங்களைத் தேடுவது, அறிவாற்றல் தேடுபொறியைப் பயன்படுத்தி (அதாவது, AI-செயல்படுத்தப்பட்ட ஒன்று). உதாரணமாக, SUVகள், ஸ்டேஷன் வேகன்கள் மற்றும் கொர்வெட்டுகள் அனைத்தும் "கார்கள்" தொடர்பானவை என்பதால், "கார்" என்ற முக்கிய சொல்லைத் தேடுவது அவற்றைப் பற்றிய ஆவணங்களைக் காண்பிக்கும்.



மொழி மொழிபெயர்ப்பு:

பல ஆண்டுகளுக்கு முன்பு, ஒரு வாக்கியத்தை ஒரு மொழியில் இருந்து மற்றொரு மொழிக்கு மொழிபெயர்ப்பது எப்போதும் குழப்பமான மற்றும்/அல்லது புண்படுத்தும் முடிவுகளை உருவாக்கும். இது மிகவும் பொதுவானது, உரை மொழிபெயர்ப்பு

சரியானதாக இருக்குமா என்று பலர் கேள்வி எழுப்பினர். AI மற்றும் NLP க்கு நன்றி, இப்போது பல மொழிகளில் எழுதப்பட்ட உரைகளில் அல்காரிதம்களைப் பயிற்றுவிப்பதன் மூலம் ஒப்பிடக்கூடிய அர்த்தத்தை வேறொரு மொழியில் உருவாக்க முடியும். ரஷியன் மற்றும் சீனம் போன்ற காதல் அல்லாத மொழிகள் கூட, பாரம்பரியமாக அவற்றின் தனித்துவமான அகரவரிசை கட்டமைப்புகள் மற்றும் எழுத்துக்களை விட சின்னங்களைப் பயன்படுத்துவதால் மொழிபெயர்ப்பது மிகவும் சவாலானது, இப்போது இந்த நுட்பத்தால் ஆதரிக்கப்படுகிறது.



NLPக்கான அணுகுமுறைகள்:

1. ஹியூரிஸ்டிக் முறை அணுகுமுறை (1950 முதல்):

சிக்கலைத் தீர்க்கும் அல்லது சுய-கண்டுபிடிப்புக்கான எந்தவொரு உத்தியும், சிறந்த, சரியான அல்லது பகுத்தறிவு என்று உறுதி செய்யப்படாத நடைமுறை முறையைப் பயன்படுத்துகிறது, இருப்பினும் உடனடி, குறுகிய கால இலக்கு அல்லது தோராயத்தை அடைவதற்குப் போதுமானது, இது ஹூரிஸ்டிக் அல்லது ஹூரிஸ்டிக் முறை என அழைக்கப்படுகிறது.

எடுத்துக்காட்டுகள்:

வழக்கமான வெளிப்பாடுகளுடன் பேட்டர்ன் பொருத்தம்

Wordnet-ஒரு வார்த்தையைப் பற்றிய தரவை மற்ற சொற்களைப் பற்றிய தரவுகளுடன் ஒழுங்கமைக்கவும்.

திறந்த மனதுடன், தர்க்கரீதியான விசாரணைகள்

நன்மைகள்:

மனித ஈடுபாட்டின் விளைவாக, அது துல்லியமானது.

இது ஒரு உடனடி உத்தி.

தீமைகள்:

பயனுள்ள ஹூரிஸ்டிக் பயன்பாட்டிற்கு அறிவும் அனுபவமும் தேவை.

நீங்கள் பல நிபுணர்களின் முடிவுகளை இணைக்க வேண்டும்.

எப்போதாவது நேரத்தை எடுத்துக்கொள்ளும்.

2. இயந்திர கற்றல் அணுகுமுறை (1990 முதல்):

நாம் ஹூரிஸ்டிக் நுட்பத்தைப் பயன்படுத்தும்போது சிக்கல் தீர்க்கப்படுகிறது, ஆனால் ஒரு திறந்த நிலை ஏற்படும் போது, ஹூரிஸ்டிக் அணுகுமுறை குறைகிறது, மேலும் இயந்திர கற்றல் அணுகுமுறை படத்தில் நுழைகிறது. NLP இன் குறிப்பிடத்தக்க வளர்ச்சி 1990 இல் தொடங்கியது, மேலும் பல புதிய உத்திகள் துறையில் வெளிப்பட்டன.

ஏற்கனவே உள்ள தரவைப் பயன்படுத்தி ஒரு மாதிரியைப் பயிற்றுவிக்கும் திறன் இயந்திர கற்றல் அணுகுமுறையைப் பயன்படுத்துவதன் முக்கிய நன்மையாகும். மேலும், இந்த அணுகுமுறை சிறந்த முடிவுகளைத் தருகிறது.

இயந்திர கற்றல் பணிப்பாய்வு

1. பிரச்சனை அறிக்கை
2. தரவு சேகரிப்பு
3. உரை முன் செயலாக்கம்
4. அம்ச பொறியியல்
5. மாதிரி கட்டிடம்
6. மாதிரி மதிப்பீடு
7. மாதிரி வரிசைப்படுத்தல்

உரையை எண்ணாக மாற்றி, பயிற்சி மற்றும் மதிப்பீட்டிற்காக இந்தத் தரவை இயந்திரக் கற்றல் மாதிரிக்கு வழங்கவும்.

பயன்படுத்தப்பட்ட மாதிரி

நைவ் பேயஸ்

லாஜிஸ்டிக் பின்னடைவு

எஸ்.வி.எம்

மறைக்கப்பட்ட மார்டோவ் மாதிரிகள்

3. ஆழ்ந்த கற்றல் அடிப்படையிலான அணுகுமுறை (2010 முதல்):

2010 முதல், ஆழமான கற்றல் தரவு அறிவியல் துறையில் பரவலாகப் பயன்படுத்தப்படுகிறது, ஏனெனில் ஆழ்ந்த கற்றல் இயந்திர கற்றலை விட நன்மைகளைக் கொண்டுள்ளது

ஆழ்ந்த கற்றலில், அம்சம் உருவாக்கம் தானாகவே உள்ளது.

கட்டமைக்கப்படாத தரவுகளுடன் நன்றாக வேலை செய்கிறது.

ஆழமான கற்றல் சுய-கற்றல் திறன்களைக் கொண்டுள்ளது.

உயர்தர முடிவுகளை வழங்கும் திறன்.

செலவு திறன். முதலியன

நன்மைகள்:

1. தக்கவைக்க உரையை அவற்றின் வரிசைத் தகவலில் எண்களாக மாற்றவும்.
2. அம்ச முக்கியத்துவம் தேவையில்லை.

ஆழமான கற்றல் மாதிரிகள்(DEEP LEARNING)

LSTM (நீண்ட குறுகிய கால நினைவகம்)

GRU (கேட்டட் ரெக்கரண்ட் யூனிட்)

சிஎன்என் (கன்வல்யூஷனல் நியூரல் நெட்வொர்க்குகள்)

மின்மாற்றிகள்

ஆட்டோஎன்கோடர், முதலியன

என்எல்பியின் போக்குகள்(2022):

குறைந்த-குறியீடு கருவிகள் பிரதானமாகச்
செல்கின்றன:

கடந்த காலத்தில், ஒரு NLP மாதிரியை உருவாக்க, துறையில் உறுதியான பின்னணி, திறந்த மூல நூலகங்களைப் பயன்படுத்துவதற்கான குறியீட்டுத் திறன் மற்றும் இயந்திர கற்றல் திறன் ஆகியவை தேவைப்பட்டன. இனி இல்லை. குறைந்த-குறியீடு/குறியீடு இல்லாத கருவிகள் சில காலமாக கிடைத்தாலும், அவை இந்த ஆண்டு மிகவும் பிரபலமாகிவிடும் என்று எதிர்பார்க்கப்படுகிறது. Monkey Learn போன்ற SaaS வழங்குநர்களுக்கு நன்றி, முன்பு தரவு விஞ்ஞானிகள் மற்றும் டெவலப்பர்களுக்கு மட்டுமே கிடைத்த NLP பணிகள் இப்போது தொழில்நுட்பம் அல்லாத நபர்களால் முடிக்கப்படலாம். MonkeyLearn இன் பாயிண்ட் அண்ட் கிளிக் மாடல் பில்டரின் காரணமாக 2022 ஆம் ஆண்டில் வணிகங்களால் NLP கருவிகளின் பயன்பாடு அதிகரிக்கும் என்று நாம் கணிக்க முடியும், இது ஒரு சில கிளிக்குகளில் உரை வகைப்பாடு அல்லது உணர்வு

பகுப்பாய்வு மாதிரிகளை உருவாக்குவது, பயிற்சி செய்வது மற்றும் ஒருங்கிணைப்பதை எளிதாக்குகிறது.

பரிமாற்ற கற்றல்:

ஒரு மாதிரி ஒரு வேலைக்காகப் பயிற்றுவிக்கப்பட்டு, பரிமாற்றக் கற்றல் எனப்படும் இயந்திரக் கற்றல் செயல்முறையைப் பயன்படுத்தி முதன்மைப் பணியுடன் தொடர்புடைய மற்றொரு செயல்பாட்டிற்காக மீண்டும் பயன்படுத்தப்படுகிறது. எனவே, தொடக்கத்திலிருந்தே ஒரு மாதிரியை உருவாக்கி பயிற்சி அளிப்பதை விட, முன் பயிற்சி பெற்ற மாதிரியை நீங்கள் மாற்றியமைக்க வேண்டும், இது விலை உயர்ந்தது, நேரத்தை எடுத்துக்கொள்வது மற்றும் அதிக அளவு தரவு தேவைப்படுகிறது. குறைந்த அளவு லேபிளிடப்பட்ட தரவைப் பயன்படுத்துவதன் மூலம், நிறுவனங்கள் என்னல்பி பணிகளை விரைவாகச் செய்யலாம். இதுவரை கணினி பார்வையில் மட்டுமே பயன்படுத்தப்பட்ட பரிமாற்றக் கற்றல், இப்போது நோக்கம் போன்ற NLP பணிகளுக்குப் பயன்படுத்தப்படுகிறது.

இந்தக் கட்டுரையிலிருந்து எடுக்கப்பட்ட

முக்கிய அம்சங்கள் –

1950 மற்றும் 1960 க்கு இடையில், இயற்கை மொழி செயலாக்க காலம் தொடங்கியது.

இயற்கையான மொழி செயலாக்கத்தால் பல பயன்பாடுகள் செய்யப்படலாம். சூழல் சார்ந்த விளம்பரங்கள், ஸ்பேம் வடிகட்டுதல், சமூக ஊடகங்கள் (வயது வந்தோருக்கான

உள்ளடக்கத்தை அகற்றுதல்), சாட்போட்கள் மற்றும் பிற எடுத்துக்காட்டுகள் ஆகியவை நாம் பார்த்தவை.

ஹூரிஸ்டிக் நுட்ப அணுகுமுறைகள் துல்லியமானவை மற்றும் விரைவானவை என்றாலும், அவை பயிற்சி மற்றும் அனுபவத்தின் மூலம் திறமையாக பயன்படுத்தப்பட வேண்டும்.

இயந்திர கற்றல் அணுகுமுறையைப் பயன்படுத்துவதன் நன்மை என்னவென்றால், ஏற்கனவே இருக்கும் தரவைப் பயன்படுத்தி ஒரு மாதிரியைப் பயிற்றுவிக்கலாம், மேலும் இந்த மாதிரி சிறந்த முடிவுகளைத் தருகிறது.

NLP தரவுகளில் பெரும்பாலானவை கட்டமைக்கப்படாதவை, மேலும் இந்த வகை தரவுகளுடன் ஆழ்ந்த கற்றல் பயனுள்ளதாக இருக்கும்.

லோ-கோட் கருவிகள், சாட்பாட்கள் மற்றும் வாடிக்கையாளர் உதவியாளர்கள் அடுத்த நிலைக்கு கொண்டு செல்லப்பட்டது, அதே போல் என்எல்பியின் புதிய மேம்பாடுகள்/போக்குகளான பரிமாற்ற கற்றல், போலி செய்திகள் மற்றும் சைபர்புல்லிங் கண்டறிதல் போன்றவை நாங்கள் கடைசியாக கவனித்த விஷயங்கள்.

