بسم الله الرّحمن الرّحیم

پژوهش درس هوش مصنوعی

برچسب گذاری بر اساس مقوله‌ی دستوری

پژوهش‌گر

علیرضا رضازاده مهریزی

۸۵۵۲۱۲۰۱

استاد

دکتر بهروز مینایی بیدگلی

پاییز و زمستان ۱۳۸۸

۱- مقدّمه

در حیطه‌ی پردازش زبان‌های طبیعی دسته‌ای از زمینه‌های کاری وجود دارند که با مقوله‌ی دستوری واژگان متنی مواجه هستند. یکی از این موضوعات " برچسب گذاری بر اساس مقوله‌ی دستوری[۱] " است. هدف از این کار پذیرش یک متن- که در اینجا از آن به پیکره[۲] یاد می‌شود - به عنوان وردی و اعمال یک سلسله تحلیل-های آماری برروی آن و سرانجام تعیین مقوله (نقش) دستوری واژگان آن به صورت برچسب‌گذاری است.

۲- مدل‌های آماری

یکی از کاربردهای رایج که برچسب گذاری برمبنای مقوله‌ی دستوری دارد پیش بینی واژه‌ی بعدی در جملات ناقص یا در حال کامل شدن است. برای نمونه از کاربردهای موضوع تشخیص واژه‌ی بعدی در متون و نیز در مباحث گفتاری مواردی از قبیل تشخیص گفتار، تشخیص دست خط، ارتباط تقویتی برای معلولان زبانی و کشف خطاهای املایی را می‌توان نام برد. منظور از کشف خطاهای املایی تصحیح خطای املایی بر اساس متن موجود است. یعنی با توجه متن مورد نظر و واژه‌های پیرامون این تشخیص صورت می‌گیرد.

برای هدف ذکر شده در بالا شاید اولین چیزی که به ذهن می‌رسد احتمال رخداد دنباله‌ای از لغات است. غالبا الگوریتم‌هایی که برای تعیین احتمال رخداد دنباله‌ای از کلمات مورد استفاده قرار می‌گیرند برای محاسبه‌ی احتمال یک واژه به عنوان واژه‌ی بعدی در یک متن ناقص نیز قابل استفاده‌اند. در مقابل از الگوریتم‌های مربوط به تشخیص کلمه‌ی بعدی برای تخصیص احتمال به دنباله‌ای از لغات نیز می‌توان بهره برد.

همان گونه که گفته شد بایستی به مطالعه‌ی احتمال بپردازیم. شاید این سوال مطرح شود که احتمال با توجه به چه داده‌ای احتمال محاسبه می‌شود؟ داده‌ی مورد نظر برای بررسی و تخصیص احتمال "پیکره" است. پیکره شامل حجم زیادی از متون متنوع یک زبان است و منبعی غنی برای مطالعات آماری به شمار می‌آید. به طور مثال پیکره‌ی Brown پیکره‌ای معروف برای زبان انگلیسی است. این پیکره که در سال‌های ۱۹۶۳ و ۱۹۶۴ ارائه شده است از روزنامه‌ها، رمان‌ها، متون علمی و متون دیگر جمع آوری شده است. مساله‌ی شمارش لغات یکی از موضوعاتی است که در ابتدای کار برچسب گذاری مورد مطالعه قرار می‌گیرد. برای این مساله دانستن چند اصطلاح رایج ضروری است.

---

۱ Part of speech tagging(POS Tag)
۲ Corpus

lema: مجموعه‌ای از فرم‌های نحوی که ریشه‌ی یکسانی دارند. مثلا دو واژه‌ی کتاب و کتاب‌ها داخل یک lema قرار می‌گیرند.

wordform: به فرم صرفی لغات در متن اطلاق می‌گردد. به عنوان مثال دو واژه‌ی کتاب و کتاب‌ها براساس wordform دارای موجودیت‌های متفاوت‌اند.

type: منظور از type کلمات متفاوت از نظر شکل ظاهری و نوشتاری است. مثلا اگر واژه‌ی کتاب چندین بار در پیکره‌ای به کار رفته باشد گوییم type همه‌ی آن‌ها یکی است.

با فرض اینکه قرار گرفتن هر کلمه در جای صحیح یک رخداد مستقل باشد و دنباله‌ای از n کلمه را به شکل

$$w_1 w_2 \dots w_n = w_1^n$$

نمایش دهیم با توجه به قانون زنجیری احتمال شرطی داریم:

$$p(w_1^n) = p(w_1)p(w_2|w_1)p(w_3|w_1^2)\dots p(w_1|w_1^{n-1})$$

محاسبه‌ی $p(w_1|w_1^{n-1})$ کار ساده‌ای نیست و نیاز به پیکره‌های بسیار بزرگ دارد. زیرا باید تعداد احتمال
دفعاتی را که یک واژه پس از رشته‌ای طولانی آمده است محاسبه کنیم.
برای کل مشکل از تخمین استفاده می‌کنیم. به این صورت که احتمال فوق را با احتمال رخداد یک واژه با توجه به واژه‌ی قبل از آن تخمین می‌زنیم. این روش بر پایه‌ی فرض مارکوف استوار است. بر اساس این فرض احتمال وقوع یک واژه تنها به واژه‌ی ماقبل آن بستگی دارد. نام دیگری که برای این روش به کار می‌رود bigram است. بر این اساس داریم:

p(کیف|خرید) = (علی دو روز پیش یک کتاب ...| خرید)p

مدل‌های مارکوفی دسته‌ای از مدل‌های احتمالی‌اند که فرض اساسی در آن‌ها این است که می‌توان برخی از ویژگی‌های مدل را بدون نیاز به بررسی گذشته بسیار دور آن پیش‌بینی نمود.

۳- زنجیرهای مارکوفی -

یک زنجیر مارکوفی، که نام آن از نام Andrey Markov اقتباس شده است، فرایند تصادفی گسسته‌ای است که دارای ویژگی مارکوفی است.گسسته بودن بدین معنا است که سیستم می‌توان در وضعیت‌های متنوعی قرار داشته باشد و تغییر وضعیت‌ها به طور تصادفی در فضایی گسسته رخ می‌دهند.

ویژگی مارکوفی بیانگر آن است که توزیع احتمال برای وضعیت آتی سیستم، به طور کلی برای همه‌ی وضعیت‌های پیش رو، تنها به وضعیت جاری سیستم وابسته و از وضعیت‌های پیشین مستقل است.

یک شرط مهم برای اینکه مجموعه‌ای از رخدادها تعریف زنجیرهای مارکوفی را برآورده کنند، استقلال آن‌ها از یکدیگر است. در زیر تعریف رسمی زنجیر مارکوفی بیان می‌شود.

زنجیر مارکوفی دنباله‌ای از متغیرهای تصادفی X_1وX_2و...وX_n است که دارای ویژگی مارکوفی تعریف شده در زیراند.

$$p(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n) = p(X_{n+1} = x \mid X_n = x_n)$$

زنجیرهای مارکوفی همگن از نظر زمانی

به زنجیر مارکوفی همگن از نظر زمانی گویند که برای همه‌ی مقادیر n احتمال گذر بین وضعیت‌ها مستقل از n باشد. این مفهوم به زبان رسمی به صورت زیر تعریف می‌شود.

$$p(X_{n+1} = x \mid X_n = y) = p(X_n = x \mid X_{n-1} = y)$$

زنجیرهای مارکوفی ثابت از نظر زمانی

به زنجیر مارکوفی ثابت از نظر زمانی گویند که اندیس‌های آن مقادیری پیوسته باشند. به عبارت دیگر متغیرهای تصادفی آن پیوسته باشند.

زنجیرهای مارکوفی از مرتبه‌ی m

یک زنجیره‌ی مارکوفی از مرتبه‌ی m به صورت زیر تعریف می‌شود.

برای $n>m$

$$p(X_n = x_n \mid X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_{n-1} = x_{n-1})$$
$$= p(X_n = x \mid X_{n-m} = x_{n-m}, X_{n-m+1} = x_{n-m+1}, X_{n-m+2} = x_{n-m+2}, \ldots, X_{n-1} = x_{n-1})$$

یعنی وضعیت بعدی به m وضعیت پیشین دارد.

چند رابطه‌ی مفید درباره‌ی زنجیرهای مارکوفی

۱) احتمال تغییر وضعیت از i به j طی n گام زمانی $p_{ij}^{(n)} = p(X_n = j \mid X_0 = i)$

۲) احتمال گذر تک گامی از i به j $p_{ij} = p(X_1 = j \mid X_0 = i)$

۳) برای یک زنجیر مارکوفی ثابت از نظر زمانی $p_{ij}^{(n)} = p(X_{n+k} = j \mid X_k = i),$

$p_{ij} = p(X_{k+1} = j \mid X_k = i)$

۴) یک گذر n گامی برای هر k که $0<k<n$ در معادله‌ای موسوم به Chapman-Kolmogorov صدق میکند.(منظور از S فضای حالت است).

$$p_{ir}^{(n)} = \sum_{r \epsilon S} p_{rj}^{(n-k)} \; p_{ij}^{(n)}$$

۳- N-gram

در بخش قبل زنجیرهای مارکوف را معرفی کرده، روابطی پیرامون آن‌ها بیان نمودیم. در بخش به مفهومی به نام N-gram می‌پردازیم.

اگر bigram را مانند یک مدل زنجیره‌ای مارکوف بدانیم، که حقیقت امر همین است، مفهوم N-gram را معادل مدل زنجیره‌ای مارکوف از مرتبه‌ی N-1 تعریف می کنیم. یعنی

$$p(w_n \mid w_1^{n-1}) = p(w_n \mid w_{n-N+1}^{n-1})$$

در زیر مثالی از یک سیستم درک گفتار مطرح می‌کنیم.

فرض کنید یک راهنمای رستوران داریم که بر مبنای گفتار عمل می‌کند. به این روش که کاربران سوالاتی پیرامون رستوران‌های موجود در شهرهای برکلی و کالیفرنیا می‌کنند. به عنوان نمونه چند مورد از محاورات به شرح زیر می‌باشند.

I’m looking for Cantonese food.
I’d like to eat dinner someplace nearby.
Tell me about Chez Pannisse.
Can you give me a listing of the kinds of food that are available?
I’m looking for a good place to eat breakfast.
I definitely do not want to have cheap Chinese food.
When is Caffe Venezia open during the day?
I don’t wanna walk more than ten minutes.

جدول زیر احتمال ظاهر شدن برخی واژه‌ها را پس از واژه‌ی eat نمایش می‌دهد.

eat on	0.16	eat Thai	0.03
eat some	0.06	eat breakfast	0.03
eat lunch	0.06	eat in	0.02
eat dinner	0.05	eat Chinese	0.02
eat at	0.04	eat Mexican	0.02
eat a	0.04	eat tomorrow	0.01
eat Indian	0.04	eat dessert	0.007
eat today	0.03	eat British	0.001

جدول ۱

همان طورکه از اعداد جدول پیدا است، پس از کلمه‌ی eat بایستی عبارت اسمی قرار گیرد.

جدول زیر نشان دهنده‌ی برخی احتمالات گرامری محاورات است. <S> به معنای آغاز جمله است.

<S> I	0.25	I want	0.32	want to	0. 65	to eat	0.26	British food	0.60
<S> I'd	0.06	I would	0.29	want a	0.05	to have	0.14	British restaurant	0.15
<S> Tell	0.04	I don't	0.08	want some	0.04	to spend	0.09	British cuisine	0.01
<S> I'm	0.02	I have	0.04	want Thai	0.01	to be	0.02	British lunch	0.01

جدول۲

با توجه به جداول بالا احتمال جمله‌ی "I want to eat British food" مطابق زیر محاسبه می‌شود.

p(I want to eat British food) =

p(I|<S>) p(want|I) p(to|want) p(eat|to) p(Britis|eat) p(food|British)

= 0.25*0.32*0.65*0.26*0.002*0.60 = 0.00016

به دلیل اینکه جمع لگاریتم‌ها معادل ضرب خطی است غالبا ابتدا لگاریتم احتمالات را با یکدیگر جمع نموده سپس از حاصل لگاریتم معکوس می‌گیریم. این روش در جاهایی که نیاز به ضرب تعداد زیادی عدد است مفید است. لازم به ذکر است که پایه‌ی لگاریتم در اکثر موارد برابر ۲ انتخاب می‌شود.

از آنجایی که یک مدل احتمالی، مثلا یک N-gram، از پیکره‌ای که مورد مطالعه است به دست می‌آید، این پیکره بایستی با دقت فراوان طراحی شود. به گونه ای که نه بیش از حد محدود به یک موضوع خاص باشد و نه بیش از اندازه عام باشد.

اگر جمله‌ای که می‌خواهیم احتمال آن را محاسبه کنیم جزئی از پیکره‌ی مورد مطالعه باشد بالطبع احتمال به دست آمده عدد بزرگتری نسبت به مقدار حقیقی خواهد بود. برای جلوگیری از این خطا پیکره را به دو بخش تقسیم می‌کنیم. بخشی از پیکره را برای بدست آوردن پارامترها و بخش دیگر را برای آزمودن پارامترها به کار می‌بریم. به بخش اول در اصطلاح مجموعه‌ی یادگیری و به بخش دوم مجموعه‌ی آزمایش می‌گویند. گاهی اوقات پیکره را به دو بخش یاد شده تقسیم می‌کنند و دو N-gram متفاوت را با استفاده از بخش یادگیری بدست می‌آورند و سپس هر دو را روی بخش آزمایش مورد تحلیل قرار می‌دهند تا ببینند کدامیک بخش آزمایش را به نحو بهتری مدل می‌کند.

۵- مدل پنهان مارکوفی

اگر سیستمی را که از فرایند مارکوفی که در آن وضعیت‌ها قابل مشاهده و دسترسی نیستند پیروی کند مدل کنبم به مدل آماری ایجاد شده مدل پنهان مارکوفی گویند.

در مدل معمولی مارکوفی وضعیت‌ها به طور مستقیم قابل رویت هستند و احتمال‌های گذر بین وضعیت‌ها تنها پارامترهای موجود هستند. هر چند در مدل پنهان مارکوفی وضعیت‌ها مستقیما قابل مشاهده نیستند، خروجی وابسته به هر وضعیت معلوم و مشخص است. هر وضعیت برای خروجی‌های خود توزیع احتمال ویژه‌ی خود را دارد. از این رو است که دنباله‌ی token هایی که یک مدل پنهان مارکوفی تولید می‌کند اطلاعاتی درباره‌ی وضعیت‌ها به دست می‌دهد.

برای مثال فرض کنید در یک اتاق پرده‌ای وجود دارد. فردی در پشت پرده مشغول آزمایش پرتاب سکّه است. فرد پشت پرده به شما نمی‌گوید که دقیقا چه می‌کند. تنها اطلاعاتی که به شما داده می‌شود نتیجه‌ی هر بار پرتاب است. مثلا

خروجی : خط خط خط خط خط خط خط خط خط خط ... خط

$$O = O_1\ O_2\ O_3\ O_4\ O_5\ O_6\ O_7\ O_8\ O_9\ O_{10}\ O_{11}O_{12}O_{13}\ \dots\ O_T$$

می‌خواهیم ببینیم چگونه می‌توان یک مدل پنهان مارکوفی ساخت که دنباله‌ی مشاهده شده از شیر و خطها را توصیف کند؟

عناصر یک مدل پنهان مارکوفی

در زیر عناصر اساسی برای تعریف یک مدل پنهان مارکوفی به طور اجمالی معرفی شده‌اند.

T: طول دنباله زمانی مشاهده یا همان تعداد پالس‌های ساعت

N: تعداد وضعیت‌های موجود در مدل

M: تعداد سمبل‌های مشاهده

Q: مجموعه وضعیت‌ها($Q = \{q_1, q_2, ..., q_N\}$).

V: مجموعه‌ی گسته‌ای از مشاهدات سمبل‌های ممکن($V = \{ v1, v2, ..., vN \}$).

A: توزیع احتمال گذار بین وضعیت‌ها $A = \{a_{ij}\}\ a_{ij} = p(q_j \text{ at } t+1 \mid q_i \text{ at } t)$

B: توزیع احتمال نماد مشاهده شده در وضعیت j $B = \{ b_j(k)\}\ b_j(k) = p(v_k \text{ at } t \mid q_j \text{ at } t)$

π: توزیع احتمال وضعیت اولیه $\pi = \{\pi_i\}\ \pi_i = p(q_i \text{ at } t = 1)$