



دانشگاه صنعتی شریف

دانشکده زبان‌ها و زبان‌شناسی

سوالات پایان‌ترم

درس پردازش زبان طبیعی

استاد: دکتر محمد بحرانی

وحید مواجی

89702855

زمستان 1390

1. در گرامرهای مستق از متن افزوده (Augmented CFGs) ویژگی‌های مختلفی برای توصیف ساختارهای دستوری زبان به کار می‌رود. در درس NLP چند ویژگی (feature) برای توصیف جملات ساده خبری در زبان انگلیسی مورد بررسی قرار گرفت. علاوه بر ویژگی‌های مطرح‌شده در درس، ویژگی‌هایی را که برای توصیف ساختارهای پیچیده‌تر دستوری (مانند جملات سوالی، جملات دارای افعال کمکی یا مجهول، جملات دارای بند موصولی (relative clause) و ...) به کار می‌روند، بررسی کنید و قواعد گرامری که این ویژگی‌ها را به کار می‌برند، بنویسید. زبان مورد نظر را می‌توانید انگلیسی، فارسی یا هر دو انتخاب کنید.

تفسیر معنایی گفتار به روابط ساختاری قسمت‌های مختلف گفتار بستگی دارد. برای ساختن گرامری برای روساخت گفتار، باید جنبه‌های مرتبط با معنای ساختار آن را در نظر گرفت. دو پدیده با اهمیت اصلی برای تفسیر معنایی بسیار زیاد به ساختار گفتار وابسته‌اند: وابستگی به بافت¹ و ساختار بیانی².

وابستگی به بافت معنای گفتار، یک پدیده فراگیر در زبان است. وقتی گفتاری را بصورت ایزوله تحلیل می‌کنیم، به دلیل پدیده‌های زیر، معنای آن به درستی تعیین نمی‌شود:

- **نمایه‌ای بودن³**. به ندرت عبارتی یا سوالی درباره کل جهان مطرح می‌شود: در هر عبارتی یک چارچوب ضمنی زمانی، فضایی و موضوعی نهفته است که حوزه‌های معنی‌شناختی مولفه‌های گفتار را تعیین می‌کند.
- **ارجاعات⁴**. بسیاری از گفتارهای شامل ضمائر و توصیفات مشخصی هستند که ارجاعات آشکاری به گفتار قبلی دارند. تحلیل ساختار گفتار قبلی برای حل درست چنین ارجاعاتی لازم است.
- **آرگومان‌های ضمنی⁵**. بسیاری از کلمات زبان طبیعی از لحاظ معنی‌شناختی اشباع‌نشده⁶ هستند و به آرگومان‌های خارجی برای تفسیر صحیح نیاز دارند. با این وجود از این تأویلات می‌توان در جملاتی استفاده نمود که مشخصاً به آرگومان‌های خود اشاره نمی‌کنند، اگر این آرگومان‌ها را نتوان از قسمت‌های قبلی گفتار بدست آورد.

“John is taller [than Peter]” “What is the speed [of John’s car]?”

¹ Context Dependence

² Rhetorical Structure

³ Indexicality

⁴ Anaphora

⁵ Implicit Arguments

⁶ Unsaturated

جوهر کار متمرکز بر ساختار بیانی گفتار این است که گوینده‌ای که گفتاری را ادا می‌کند، ممکن است از کارگفتارهایی⁷ استفاده کند که نیروی منظوری⁸ آنها حوزه‌ای در عبارات پیچیده‌ای داشته باشد که از معنی جملات منفرد و روابط بیانی بین آنها تشکیل شده‌اند. این روابط بیانی ممکن است آشکارا بیان شده باشند یا ممکن است براساس معنی جمله حذف شده باشند. مثال‌های کاربردشناختی چنین عبارت پیچیده‌ای بدینقرار می‌باشند:

"A caused B", "A was caused by B", "A provides evidence for B"

که A و B ممکن است عباراتی باشند که توسط جملات منفرد ادا شده‌اند یا ممکن است خودشان آن عبارات پیچیده‌ای باشند که توسط بخش‌های گفتار بیان شده‌اند. بدلیل حالت دوم، یک تحلیل درست گفتار لازم است تا آرگومان‌های روابط بیانی را برقرار سازد. از طرفی دیگر، خود روابط بیانی یک مولفه ساختار ساز مهم برای گفتار می‌باشند.

هر جمله ابتدائاً به یک طریق محلی و مستقل از متن تفسیر می‌شود. این امر به یک بازنمایی معنایی می‌انجامد که معمولاً شامل متغیرهای آزادی می‌شود که متناظر مولفه‌های وابسته به گفتار در معنا می‌باشند. وقتی جمله‌ای را با گفتار شامل آن تحلیل می‌کنیم، این متغیرها به مقادیری که از بافت جمله استخراج می‌شوند، مقید خواهند شد.

گفتارها ساختار سلسله‌مراتبی دارند. آنها بصورت بازگشتی از واحدهایی از انواع مختلف تشکیل می‌شوند که می‌توانند به صورت سازه‌هایی⁹ از یکدیگر بیایند. برای در نظر گرفتن این مسأله، یک گرامر گفتار باید قادر باشد یک ساختار درختی به یک گفتار نسبت دهد. به این ساختار درختی، درخت پارس گفتار گفته می‌شود. برای اینکه بصورت فرمال شرح دهیم چگونه یک درخت پارس گفتار از جملات سازنده آن تشکیل می‌شود، از یک گرامر مستقل از متن استفاده می‌کنیم که نمادهای غیرپایانی آن با زوج‌های ویژگی/مقدار¹⁰ تقویت¹¹ شده‌اند. (مقوله‌های مجزای غیرپایانی، مجموعه ویژگی‌های مجزایی دارند). قواعد مستقل از متن معین می‌کنند که چگونه قسمت‌های سازنده یک گفتار (که آنها را واحدهای سازنده گفتار¹² یا dcu می‌نامیم) از روی زیرسازه‌های خود ساخته می‌شوند. مقادیر یک ویژگی در یک گره غیرپایانی نمایانگر ویژگی‌های ساختاری و معنایی مرتبط dcu ساخته‌شده توسط گره غیرپایانی می‌باشند. یک ویژگی، یک مجموعه ثابت از عبارات مقداری ممکن دارد. عبارات مقداری ممکن است

⁷ Speech Acts

⁸ Illocutionary Force

⁹ Constituent

¹⁰ Attribute/Value

¹¹ Augmented

¹² Discourse Constituent Unit

از انواع متفاوتی باشند: ممکن است اتمیک باشند، ممکن است خودشان مجموعه‌هایی از زوج ویژگی/مقدار باشند یا ممکن است عبارات منطقی باشند.

عبارات مقداری اغلب حاوی اطلاعات جزئی¹³ هستند. لذا ممکن است شامل متغیرهای آزاد باشم. یک عبارت مقداری نمایانگر مجموعه‌ای از نمونه‌های اصلی¹⁴ خود باشد. (لذا یک عبارت مقداری بدون متغیر نمایانگر یک مجموعه منحصر بفرد¹⁵ است). وقتی یک ویژگی، یک عبارت مقداری دارد که نمایانگر مجموعه تهی می‌باشد، نماد مقوله مرکبی که این عبارت مقداری را شامل می‌شود، نمی‌تواند یک dcu ممکن را برچسب بزند.

قواعد مستقل از متن با نظر به این حقیقت که رخدادهای متفاوت یک متغیر یکسان، مقادیر یکسانی را به خود می‌گیرد، تطابق¹⁶ و وراثت به طرف بالا¹⁷ی ویژگی‌های مرتبط سازه‌های مختلف را تحمیل می‌کند. برای اینکه این مسأله را دقیقتر واکاوی کنیم، معنی یک قاعده مستقل از متن تقویت شده را بدین صورت تعریف می‌کنیم: اگر A, B, C, Y, Z نمایانگر نمادهای مقوله‌ای پیچیده باشند (شامل زوج‌های ویژگی/عبارت-مقداری) و $[dcu1]_y$ و $[dcu2]_z$ ، dcuهای معتبری باشند، آنگاه قاعده " $A \Rightarrow B \ C$ "، $\delta([[dcu1]_y [dcu2]_z]_A) \circ dcu$ را معتبر می‌سازد اگر و فقط اگر تعویض δ ، عمومی‌ترین یکی‌کننده¹⁸ عبارات $\langle B, C \rangle$ و $\langle Y, Z \rangle$ باشد و $\delta(A)$ یک نماد مقوله‌ای مرکب معتبر باشد که شامل عبارات ویژگی-مقدار نباشد. انتشار مقادیر ویژگی‌ها بین dcuها نقش مهمی در تفسیر معنایی عبارات گفتار ایفا می‌کند. اکنون برخی از این ویژگی‌ها را نام برده و نقش آنها را در تفسیر معنایی و پردازش گفتار شرح می‌دهیم.

ویژگی معناسناسی¹⁹ حاوی یک فرمول منطقی است که نمایانگر معنی واحد سازنده گفتار متناظر خود می‌باشد. عناصر ارجاعی که درون dcu مشخص نشده‌اند، با متغیرهای آزاد نشان داده می‌شوند. دو راهکار برای تشخیصی ارجاع²⁰ وجود دارد: (1) یکی‌شدن با عبارات-مقداری ویژگی‌های دیگر dcuها و (2) جستجوی صریح در مجموعه مدلول‌های²¹ گفتار dcuهای موجود.

¹³ Partial

¹⁴ Ground Instances

¹⁵ Singleton

¹⁶ Agreement

¹⁷ Upwards-Inheritance

¹⁸ Most General Unifier

¹⁹ Semantics

²⁰ Anaphor Resolution

²¹ Referent

مجموعه مدلول‌های گفتار، موجودیت‌هایی که در واحد گفتار وجود دارند را شامل می‌شود. اینها بعلاوه موجودیت‌های درون مجموعه‌های مدلول‌های گفتار واحدهای دربرگیرنده (گره‌های بالایی در درخت)، موجودیت‌هایی هستند که در دسترس ارجاعات یک گفتار می‌باشند. آن واحد گفتاری را توسعه و گسترش می‌بخشند. هر مدلول گفتار یک زوج است که شامل این موارد می‌شود: (1) عبارت زبان‌شناختی که آنرا مطرح کرده و (2) بازنمایی معناشناختی که سیستم به آن عبارت ملحق کرده است. مجموعه‌های مدلول گفتار با الگوریتم‌های مختلفی بررسی می‌شوند که معنی انواع مختلفی از عبارات وابسته به متن را بدست می‌آورند مانند توضیحات معین، ضمائر ارجاعی، ضمائر اشاری، کلمات "one" و "one" و آرگومان‌های ضمنی اسامی تابعی، مقایسه‌ای و غیره.

زمان-ارجاع²² بازه زمانی را شامل می‌شود که بعنوان یک اندیس زمانی برای شرح و بسط واحد گفتاری مورد نظر عمل می‌کند. زمان-ارجاع را می‌توان با رخداد قیده‌های زمانی صریح، بازنشانی²³ نمود. روایات²⁴، نوع خاصی از واحدهای گفتاری هستند که زمان در آنها نقش خاصی ایفا می‌کند؛ آنها با رخداد جملاتی که جنبه غیر استمرار²⁵ (جملات رویدادی²⁶) دارند مشخص می‌شوند. در یک روایت، زمان-ارجاع هرگاه که رویدادی رخ می‌دهد، بازنشانی می‌شود.

نمایه فضایی²⁷ و نمایه وجه‌نما²⁸، در تفسیر معنای گفتارها نقشی مشابه با زمان-ارجاع دارند: آنها مقداری برای مولفه‌های نمایه ارزیابی فرمول معنی‌شناختی مشخص می‌کنند. در زبان انگلیسی، آنها فقط با رخداد قیده‌های فضایی و وجه‌نما یا دیگر اطلاعات صریح مشابه روزرسانی می‌شوند.

قواعد گرامری

پدیده‌های معنایی که به آنها توجه می‌کنیم شامل تشخیص ارجاعات، حوزه نمایه‌های وجه‌نما، حرکت زمان ارجاع و روابط بیانی هستند. گرامر از قواعدی تشکیل می‌شود که شرح می‌دهند چگونه انواع مختلفی از واحدهای سازنده با ساختار متفاوت را باید بنا کرد. بین انواع dcu‌های زیر تمییز قائل می‌شویم:

²² Reference-Time

²³ Reset

²⁴ Narratives

²⁵ Non Durative

²⁶ Event Sentences

²⁷ Temporal Index

²⁸ Modal Index

• **پیروها²⁹**. اینها ساختارهای دودویی هستند که مولفه اول آنها در دسترسی باقی می‌ماند؛ آنها را بصورت واحدهایی در نظر می‌گیریم که همه یا اغلب ویژگی‌های مرتبط ساختاریشان از سازه سمت چپ به ارث رسیده‌اند. (در گفتار، برخلاف جمله، عنصر پایه همیشه در سمت چپ عنصر پیرو قرار می‌گیرد). در پیروهای معنایی، یک رابطه معنایی بین دو سازه وجود دارد. این حالت در پیروهای بیانی و زنجیره‌های موضوع-حاکم³⁰ وجود دارد. وقفه‌ها³¹، از طرفی دیگر، با اینکه از لحاظ ساختاری مشابهند، از لحاظ معنایی بسیار متفاوت می‌باشند: در این حالت به هیچ وجه رابطه‌ای معنایی بین دو سازه وجود ندارد.

• **هماهنگی‌های دودویی³²**. اینها ساختارهایی دودویی هستند که مولفه دوم آنها وضعیت مشابهی با مولفه اول دارد، لذا مولفه اول را غیرقابل دسترس می‌سازد. تحت این مقوله، هماهنگی‌های بیانی³³ (همتای پیروهای بیانی) و زوج‌های مجاورت³⁴ که مرتبط با بعد تعاملی گفتار هستند مطرح می‌شوند (شامل زوج‌های سوال/جواب و درخواست/پاسخ).

• **هماهنگی‌های N تایی³⁵**. اینها ساختاری مسطح هستند که می‌توانند شامل عناصر متعدد دلخواهی باشند که در هر زمان؛ متأخرترین آنها در دسترس است. اینها شامل لیست‌ها، لیست‌های یکنواخت³⁶ و روایات می‌شوند. برای ساختن هماهنگی‌های N تایی با استفاده از گرامر مستقل از متن، باید به آنها یک ساختار بازگشتی نسبت دهیم: آنها را با استفاده از قواعد دودویی که به سمت راست گسترش می‌یابند، می‌سازیم.

نمادهای مقوله‌ای بصورت $cat[\varphi_1:\alpha_1, \dots, \varphi_n:\alpha_n]$ می‌باشند که “cat” نماد غیر پایانی پایه گرامر مستقل از متن می‌باشد، $\varphi_1, \dots, \varphi_n$ ویژگی‌ها می‌باشند و $\alpha_1, \dots, \alpha_n$ عباراتی هستند که نمایانگر مجموعه مقادیر این ویژگی‌ها می‌باشند. متغیرها بصورت *یتا/لیک* نشان داده می‌شوند. اگر در یک قاعده ساخت گروهی³⁷، یک ویژگی، بعنوان عبارت-مقداری‌اش، متغیری داشته باشد که هیچ جای دیگر در قاعده وجود ندارد، برای خوانایی بیشتر، آن ویژگی را در مشخصات آن قاعده نشان نمی‌دهیم.

²⁹ Subordination

³⁰ Topic-Dominant Chains

³¹ Interruptions

³² Binary Coordinations

³³ Rhetorical Coordinations

³⁴ Adjacency Pairs

³⁵ N-ary Coordinations

³⁶ Monotonic

³⁷ Phrase Structure Rule

گرامر، نماد آغازین خاصی ندارد؛ dcu های همه انواع را بصورت گفتارهایی خوش ساخت در نظر می گیریم. جملات را بعنوان dcu های پایه فرض می کنیم. قواعد زیر ساختار داخلی برخی از مهمترین انواع واحدهای سازنده گفتار مرکب را مشخص می سازند.

Lists:

List [drs: $d1 \cup d2$, schema:s, sem:s(x) & s(y)]

=> $dcu1$ [drs:d1, sem:s(x)]

$dcu2$ [drs:d2.sem:s(y)]

ایده ساده پشت این قاعده این است که اگر دو dcu مجاور را بتوان با این فرض که ساختارهای معنای موازی دارند تحلیل کرد، آنگاه آنها را می توان در یک ساختار لیست با هم قرین ساخت. باید توجه داشت که هیچ شرطی روی مقوله dcu لحاظ نشده است: نمادهای مقوله ای اصلی در سمت راست قاعده، متغیر هستند.

شرط فرمال اعمال این قاعده، وجود یک فرمول- λ مانند s است به نحوی که محتوی معنایی هر کدام از دو dcu به شکل $s(u)$ باشند. نیاز است که بازنمایی معنای dcu ساختار $s(u)$ را داشته باشد وگرنه باید با تعداد محدودی از تبدیلات منطقی معادل، آن را بدین شکل درآورد. این محدودیت، مقادیر بدیهی s مانند را ($\lambda u: \text{if SEM1 then SEM1 else SEM2}$) مستثنی می کند. نوع متغیرهای- λ s باید عناصری در یک سلسله مراتب از پیش تعریف شده “انواع طبیعی” باشند که معنی مداخل واژگانی را طبقه بندی می کند. مقدار s که نمایانگر مخرج مشترک بین معانی دو سازه است، بعنوان مقدار ویژگی “schema” در لیست dcu ذخیره می شود. باید توجه داشت که این قاعده اغلب باعث ابهام می شود. برای مثال: اگر “John likes Mary” قبل از “Peter likes Mary” بیاید، می توان شمایهای مختلفی را تصور کرد حتی بدون در نظر گرفتن انتخاب های متفاوت برای انواع متغیرهای- λ . $\lambda P:P(\text{MARY})$ یا

$\lambda R,x:R(x,\text{MARY})$ یا

$\lambda x:\text{LIKE}(x, \text{MARY})$ یا $\lambda x,y:\text{LIKE}(x,y)$. آخرین مورد ارجح است، چون که ویژگی تر است. به جای اینکه قاعده گرامری را محدود کنیم تا فقط به نحو خاصی به هر حالت اعمال شود، فرض می شود که یک ترجیح برای اعمال ویژه ترین قاعده بصورت یک اصل شهودی در فرایند پارس اعمال می شود.

ویژگی “sem” معنای یک dcu را ذخیره می کند. عمل عملگر “&” که برای ساختن معنی بکار رفته است، عمومی تر از همتای منطقی خود است: آرگومان های آن می توانند گوینده های

مختلفی و عملگرهای نیروی منظوری متفاوتی داشته باشند. بازنمایی معنایی با دخیره "&" از محتوای گزاره‌ای گفتارهای منفرد در یک ساختار دارای داده به شدت اندیس شده ساخته می‌شود. در فرمول بکاررفته برای مقادر ویژگی "sem" یک dcu، متغیرهای بدون نوع برای نمایش ارجاعات نامشخص بکار رفته‌اند، مانند ضمایر و توصیفات معین. فرایند یکی‌سازی که ویژگی "sem" دو dcu را با هم مطابقت می‌دهد، وقتی که آنها برای ساختن یک لیست به هم متصل می‌شوند، ممکن است به جای این متغیرها، عباراتی را جایگزین کند و لذا ارجاعات را مشخص نماید. بنابراین برخی از ترجیحات قدرتمند برای مشخص کردن ارجاعات، بصورت تبعیت مستقیم از ساختار موازی، تعریف می‌شوند. در این حالت عناصر ارجاعی بدون هیچ جستجویی درون فضای مدلول‌های گفتار، مشخص می‌شوند.

برای جمع‌آوری کاندیداهای مناسب برای فرایندهای تشخیص ارجاع بعدی، مدلول‌های گفتار (مقادیر ویژگی "drs") برای هر دو dcu به شکل مجموعه مدلول‌های لیست dcu درمی‌آیند. برای بسط یک لیست dcu به لیستی که یک عنصر بیشتر داشته باشد، قاعده زیر را بکار می‌بریم:

list [schema:s, drs:drs1 \cup drs2, sem:p & s(y)]

=> list [schema:s, drs:drs1, sem:p]

Dcu[drs:drs2, sem: s(y)]

یک لیست را می‌توان به یک dcu دیگر بسط داد اگر این dcu ساختار تعریف شده در ویژگی schema لیست را مقداردهی کند.

Monotonic Lists:

m-list [schema:s, drs:drs1 \cup drs2,

direction: (if $x < y$ then incr

else if $x > y$ then dect else fail),

last: y, sem:s(x) & s(y)]

=> dcu [drs:drs1, sem:s(x)]

dcu [drs:drs2, sem:s(y)]

اگر در یک لیست، آرگومان‌های مختلف تابع schema عناصر یک حوزه خطی باشند، آن لیست احتمالاً یک لیست یکنواخت خواهد بود. برای اینکه تعیین اینکه آیا dcu بعدی را می‌توان به یک لیست یکنواخت اضافه نمود، امکان‌پذیر باشد، این لیست‌ها باید مقدار متأخرترین

آرگومان (ویژگی “last”) و “direction” (“increasing” یا “decreasing”) را در خود داشته باشند. (“fail”) یک ثابت خاص است که نمی‌توان از آن بعنوان یک عبارت-مقداری ویژگی استفاده نمود).

m-list [schema:s, drs: $drs1 \cup drs2$,

direction: (if $x < y \wedge p = \text{decr}$ then decr else fail),

last:y, sem:q & s(y)]

=> m-list [schema:s, drs:drs1, incr:p, last:x, sem:q]

Dcu [drs:drs2, sem:s(y) $\wedge y \in d$]

یک لیست یکنواخت می‌تواند توسط یک dcu دیگر که ساختار تعریف‌شده توسط ویژگی schema خود را مقداردهی می‌کند، بسط داده شود با این فرض که ترتیب افزایشی یا کاهش‌ی رعایت می‌شود.

از “روایات” برای بیان یک سری از اتفاقات در نقاط پشت سر هم در یک خط زمانی اتفاق می‌افتند استفاده می‌شود. اینکه چگونه عبارت پایه³⁸ بعدی روایت با زمان مرجع قبلی آن روایت در تعامل است، به وجه³⁹ آن بستگی دارد: عبارات استمراری⁴⁰ متفاوت از عبارات غیراستمراری عمل می‌کنند. لذا به دو قاعده برای بسط یک روایت نیاز داریم. ابتدا قاعده استمراری‌ها را می‌آوریم:

narrative [drs:drs1 \cup drs2, reference-time:rt,

tense: (if $t1 \in xt$ then t1 else fail), sem: p & [s]_{rt}]

=> narrative [drs:drs1, reference-time:rt, tense:t1, sem:p]

Dcu[drs:drs2, reference-time:rt, x-tenses:xt,

Aspect:durative, sem:s]

ویژگی reference-time به بازه زمانی اشاره دارد که در آن پیشرفت زمان روایت انجام شده است. یعنی بازه زمانی بعد از آخرین رخداد در روایت تا کنون. یک dcu استمراری که روایتی را بسط می‌دهد در زمان-مرجع آن روایت ارزیابی می‌شود.

مقدار ویژگی “tense” در یک dcu روایتی، نشانگر حالت‌های متمایز زمانی مجاز روایت در یک زبان خاص است. برای زبان انگلیسی، این مسأله شامل تمایز بین گذشته، حال، ماضی بعید و آینده است. در یک جمله (یا یک dcu پیچیده‌تر) که قرار است درون یک روایت قرار

³⁸ Main Clause

³⁹ Aspect

⁴⁰ Durative

گیرد، ویژگی مرتبط، “tense” نیست بلکه “x-tenses” است که زمانهایی را ذخیره می‌کند که ممکن است از بیرون به dcu تحمیل شوند. یک جمله که زمان آن PRESENT است می‌تواند در بافتی بکار رود که زمان آن PRESENT، PAST یا FUTURE باشد. مشابهاً یک جمله با زمان PAST با یک چارچوب PAST یا PLUPERFECT سازگار می‌باشد. یک جمله با زمان PLUPREFECT فقط با یک چارچوب PLUPERFECT سازگار است. وقتی یک dcu یک روایت را بسط می‌دهد، ویژگی “tense” آن روایت باید عنصری از عناصر “x-tenses” آن dcu باشد.

narrative [drs:drs1 ∪ drs2, reference-time:v,

tense: (if $t1 \in xt$ then $t1$ else fail),

sem: $p \& [s]_u \& t <_i u <_i v$]

=> narrative [drs:drs1, reference-time:t, tense:t1, sem:p]

dcu[drs:drs2, reference-time:u,

aspect:event, x-tenses:xt, sem:s]

یک event-dcu که روایتی را بسط می‌دهد در یک بازه u بعد از زمان-مرجع جاری t روایت ارزیابی می‌شود. زمان-مرجع جدید روایت بسطیافته بازه دیگر v بعد از u است. $(a <_i b)$ یعنی a بلافاصله قبل از b می‌آید). تأثیر این امر این است که یک dcu استمراری بدون قیدهای زمانی صریح با رخداد بعدی بسته خواهد شد. تأثیر بعدی این است که رخدادهای بعدی همیشه با یک فاصله-زمانی⁴¹ جدا خواهند بود.

Rhetorical Subordinations

dcu1[$\varphi_1:\alpha_1, \dots, \varphi_n:\alpha_n$, index: i , sem: $a \& R(a, [b]_i)$]

=> dcu1 [$\varphi_1:\alpha_1, \dots, \varphi_n:\alpha_n$, index: i , sem: a]

dcu2[sem: $\lambda x:R(x,b)$]

(pop-marker)

این قاعده، پیروهای معنایی را پارس می‌کند که شامل یک رابطه بیانی پیروی صریح مثل R می‌باشند (“for instance”, “because”). فرض می‌شود معنی این رابطه در معنای dcu پیرو قرار می‌گیرد. لذا این dcu بعنوان مقداری از ویژگی “sem” خود، یک تابع λ دارد که

⁴¹ Time-Gap

به یک آرگومان گزاره‌ای نیاز دارد. ویژگی‌ها و مقادیر یک عمل پیرو بودن از سازه پایه به ارث می‌رسد.

بعد از سازه پیرو، می‌توانند یک “pop-marker” بیایند مثل “so” یا “anyway”. همه کلمات کلیدی بعنوان واحدهای مستقل و جدا از جملاتی که قبل یا بعد از آنها می‌آیند، دیده می‌شوند. در فرمول‌بندی قاعده پیروسازی معنایی یک ویژگی را فرض گرفتیم که “index” نامیده می‌شود، و شامل زمان مرجع و نمایه‌های فضایی و وجه‌نما است. این قاعده نشان می‌دهد که چگونه واحد سازنده گفتار پیرو از لحاظ معنی‌شناختی توسط واحد پایه بافت پیدا می‌کند.

$$dcu1[\varphi_1:\alpha_1, \dots, \varphi_n:\alpha_n, \text{index}:i, \text{sem}:a \ \& \ R(a,[b]_i)]$$

$$\Rightarrow dcu1(\varphi_1:\alpha_1, \dots, \varphi_n:\alpha_n, \text{index}:i, \text{sem}:a)$$

(push-marker)

$$dcu2[\text{sem}:b]$$

(pop-marker)

این قاعده پیروسازی‌های معنایی که رابطه بیانی آنها آشکارا نشان‌دار نیست را پارس می‌کند. بازه متغیر R روی همه روابط بیانی پایه است. بعد از سازه پیرو ممکن است یک push-marker مثل “like” بیاید و بعد از آن ممکن است یک pop-marker بیاید.

Rhetorical Coordinations:

$$dcu[\varphi_1:\text{mscg}(\alpha_1,\beta_1), \dots, \varphi_n:\text{mscg}(\alpha_n,\beta_n), \text{sem}:a \ \& \ b \ \& \ R(a,b)]$$

$$\Rightarrow dcu1 [\varphi_1:\alpha_1, \dots, \varphi_n:\alpha_n, \text{sem}:a]$$

$$dcu2 [\varphi_1:\beta_1, \dots, \varphi_n:\beta_n, \text{sem}:\lambda x:R(x,b)]$$

این قاعده، هماهنگی‌های معنایی را پارس می‌کند که شامل یک رابطه بیانی هماهنگی دودویی صریح می‌باشند (“therefore”, “thus”, “accordingly”). <Ref, Mann, Talmy>. مانند حالت پیروسازی که توضیح داده شد، معنی رابطه در معنی‌شناختی عبارتی که در آن قرار دارد مستتر است که علی‌القاعده نشانگر یک محمول در گزاره است/ تابع “mscg”، “خصوص‌ترین تعمیم مشترک”⁴² آرگومان‌های خود را در سلسله‌مراتب عبارات-مقداری ویژگی‌های مرتبط محاسبه می‌کند.

⁴² Most Specific Common Generalization

$$dcu[\varphi_1:mscg(\alpha_1,\beta_1),\dots,\varphi_n:mscg(\alpha_n,\beta_n), sem:a \& b \& R(a,b)]$$

$$\Rightarrow dcu1 [\varphi_1:\alpha_1, \dots, \varphi_n:\alpha_n, sem:a]$$

$$dcu2 [\varphi_1:\beta_1, \dots, \varphi_n:\beta_n, sem:b]$$

این قاعده هماهنگی‌های معنایی دودویی را پارس می‌کند که آشکارا بدان صورت نشان‌دار نشده‌اند. لذا معنی‌شناسی دومین dcu به جای اینکه یک محمول روی گزاره‌ها باشد، یک گزاره است. بازه متغیر R روی همه روابط بیانی هماهنگی دودویی است.

Topic-Dominant Chaining

$$dcu1[\varphi_1:\alpha_1,\dots,\varphi_n:\alpha_n, index:i, sem:s1(y)(x)\&[s2(y)]_i]$$

$$\Rightarrow dcu1 [\varphi_1:\alpha_1, \dots, \varphi_n:\alpha_n, index:i, sem:s1(y)(x)]$$

$$dcu2 [\varphi_1:\beta_1, \dots, \varphi_n:\beta_n, sem: s2(y)]$$

(pop-marker)

زنجیره‌های موضوع-حاکم، ساختارهای پیروسازی هستند. در این ساختارها، dcu پیرو، درباره یک سازه محمول در معنی‌شناسی dcu پایه اطلاعاتی می‌دهد. قاعده بالا مستلزم این است که عنصری مانند y وجود داشته باشد به نحوی که محمول dcu سمت چپ و معنی‌شناسی dcu سمت راست را بتوان بصورت عباراتی با ساختار $f(y)$ فرموله کرد – که مانند قبل بتوان با تعداد محدودی تبدیلات منطقی، از روی فرمولی که مسقیماً به رو ساخت dcu مربوط می‌شود، به این فرمول‌بندی رسید.

Adjacency Pairs:

$$QA [sem: a(b)]$$

$$\Rightarrow dcu1 [mood: interroogative, sem: b]$$

$$dcu2 [sem: a(b)]$$

(pop-marker)

این قاعده، زوج‌های سوال/جواب را پارس می‌کند. فرض می‌شود معنی‌شناسی یک سوال بله/خیر، یک گزاره باشد. معنی‌شناسی یک سوال wh یک عبارت set -denoting است. معنی‌شناسی یک جواب، اسنادی روی معنی‌شناسی سوال است.

$$RR [sem: a(b)]$$

=> $dcu1[\text{mood:request, speaker:p1, addressee:p2, sem:b}]$

$dcu2[\text{speaker:p2, addressee:p1, sem: } a(b)]$

(pop-marker)

این قاعده، زوج‌های درخواست/پاسخ را پارس می‌کند. از لحاظ معنی‌شناسی، اینها خیلی مشابه زوج‌های سوال/جواب هستند.

Interruptions:

$dcu1[\alpha]$

=> $dcu1[\alpha]$

(interruption-marker)

$dcu2[\beta]$

(pop-marker)

این قاعده، وقفه‌های از لحاظ معنی‌شناختی غیرمرتبط یک گفتار را مجاز می‌شمارد. وقفه‌ها ممکن است توسط نشانگرهایی مانند "Oh!" مطرح شوند.

2. یکی از کاربردهای مهم مدل مخفی مارکوف (HMM)، استفاده از آن در بازشناسی گفتار می‌باشد. در مورد روش‌ها و نحوه به کارگیری HMM در بازشناسی گفتار، با جزئیات توضیح دهید.

اجزای اصلی یک سیستم تشخیص گفتار پیوسته با واژگان گسترده در شکل 2.1 نشان داده شده است. موج صوتی ورودی از یک میکروفون طی فرایندی که استخراج ویژگی⁴³ نام دارد، به دنباله‌ای از بردارهای صوتی با طول ثابت تبدیل می‌شود. این بردارها را بصورت زیر نشان می‌دهیم:

$$Y_{1:T} = y_1, \dots, y_T$$

سپس دیکودر سعی می‌کند تا دنباله‌ای از کلمات $w_{1:L} = w_1, \dots, w_L$ پیدا کنید به نحوی که با بیشترین احتمال، Y را تولید کرده باشند، یعنی دیکودر باید این را پیدا کند:

$$\hat{w} = \arg \max_w \{P(w|Y)\} \quad (2.1)$$

هرچند از آنجایی که مدل کردن مستقیم $P(w|Y)$ دشوار می‌باشد، از قاعده بیز استفاده می‌شود تا معادله 2.1 را به مسأله معادل زیر تبدیل کند:

$$\hat{w} = \arg \max_w \{P(Y|w)P(w)\} \quad (2.2)$$

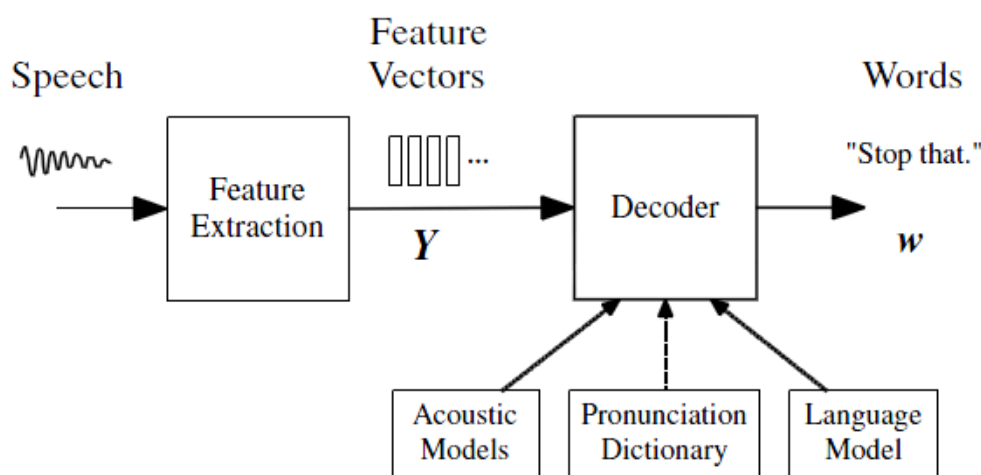
احتمال $P(Y|w)$ با یک مدل آکوستیکی⁴⁴ و احتمال $P(w)$ با یک مدل زبانی⁴⁵ تعیین می‌شوند. واحد پایه صدا در مدل آکوستیکی، آوا⁴⁶ است. مثلاً کلمه bat از سه آوا تشکیل شده است: /b/ /ae/ /t/ برای زبان انگلیسی حدود 30 آوا لازم است.

⁴³ Feature Extraction

⁴⁴ Acoustic Model

⁴⁵ Language Model

⁴⁶ Phone



شکل 2.1: معماری یک سیستم تشخیص گفتار مبتنی بر HMM

برای هر w ، مدل آکوستیکی متناظر آن با به هم پیوستن مدل‌های آوایی برای تشکیل کلماتی که توسط یک لغتنامه تلفظی تعریف می‌شوند، به وجود می‌آید. پارامترهای این مدل‌های آوایی از داده آموزشی که از موج‌های گفتار و بازنویسی آوایی آنها تشکیل می‌شود، تخمین زده می‌شوند. مدل زبانی، معمولاً یک مدل N-gram است که طبق آن احتمال هر کلمه روی احتمال N-1 کلمه قبل از خود مشروط می‌شود. پارامترهای N-gram با شمارش N-تایی‌ها در پیکره‌های متنی متناسب تخمین زده می‌شوند. دیکودر بدین گونه عمل می‌کند که در همه دنباله‌های ممکن کلمات با استفاده از هرس کردن⁴⁷ جستجو می‌کند و فرضیه‌های غیرمحمّل را کنار می‌گذارد تا جستجو را قابل کنترل نگه دارد. وقتی به انتهای گفتار رسید، محتمل‌ترین دنباله کلمات به عنوان خروجی داده می‌شود. دیکودرهای جدیدتر می‌توانند شبکه‌هایی⁴⁸ تولید کنند که بازنمایی فشرده‌ای از محتمل‌ترین فرضیه‌ها را در خود داشته باشند.

استخراج ویژگی

مرحله استخراج ویژگی سعی بر آن دارد که بازنمایی فشرده‌ای از موج گفتار ارائه دهد. این شکل باید کمترین میزان از دست دادن⁴⁹ اطلاعاتی که ممیز کلمات هستند را داشته باشد و تناظر خوبی با فرضیات توزیعی مدل آکوستیکی داشته باشد. مثلاً اگر از توزیع‌های گوسی با

⁴⁷ Pruning

⁴⁸ Lattices

⁴⁹ Loss

کوواریانس قطری برای توزیع‌های خروجی-حالت⁵⁰ استفاده می‌شود، آنگاه ویژگی‌ها باید طوری طراحی شوند که گوسی و غیرهمبسته⁵¹ باشند.

بردارهای ویژگی معمولاً هر 10 میلی‌ثانیه یکبار با استفاده از یک پنجره تحلیل همپوشان⁵² با طول تقریباً 25 میلی‌ثانیه محاسبه می‌شوند. یکی از ساده‌ترین و پرکاربردترین روش‌های کدکردن براساس روش MFCCs⁵³ کار می‌کند. این روش با اعمال یک تبدیل کسینوسی گسسته (DCT) بر روی یک تخمین طیفی لگاریتمی⁵⁴ کار می‌کند که با هموارسازی⁵⁵ یک FFT با 20 فاصله فرکانسی که بطور غیرخطی در طیف گفتار توزیع شده‌اند، محاسبه شده است. مقیاس فرکانسی غیرخطی بکار رفته، مقیاس مل⁵⁶ نام دارد و پاسخ گوش انسان را تقریب می‌زند. از DCT برای این استفاده می‌شود تا تخمین طیفی را هموار سازد و مولفه‌های ویژگی را غیرهمبسته سازد. بعد از تبدیل کسینوسی، اولین مولفه نمایانگر میانگین انرژی-لگاریتمی⁵⁷ فواصل فرکانسی است. ممکن است گاهی آن را با انرژی-لگاریتمی فریم⁵⁸ تعویض کنیم یا به کلی حذف کنیم.

محدودیت‌های روان‌آکوستیکی⁵⁹ در یک کدینگ مرتبط که پیش‌بینی خطی ادراکی⁶⁰ یا PLP نامیده می‌شود، قرار می‌گیرند. PLP ضرایب پیش‌بینی خطی را از روی یک طیفی توانی فشرده غیرخطی وزندار محاسبه می‌کند و سپس ضرایب پیش‌بینی خطی را به ضرایب کمپسترال⁶¹ تبدیل می‌کند. در عمل، PLP قادر است بهبودهای کوچکی در MFCCs مخصوصاً در محیط‌های نویزی ایجاد کند، و لذا کدینگ ارجح برای بسیاری از سیستم‌ها می‌باشد.

علاوه بر ضرایب طیفی، ضرایب رگرسیون مرتبه اول (دلتا) و مرتبه دوم (دلتا-دلتا) نیز معمولاً با یک روش هیوریستیک اضافه می‌شوند تا فرض استقلال شرطی موجود در مدل‌های آکوستیکی HMM را جبران کنند. اگر بردار ویژگی اصلی (استاتیک) y_t^s باشد، آنگاه پارامتر دلتا، Δy_t^s توسط فرمول زیر محاسبه می‌شود:

⁵⁰ State-Output

⁵¹ Uncorrelated

⁵² Overlapping Analysis Window

⁵³ Mel-Frequency Cepstral Coefficients

⁵⁴ Log Spectral Estimate

⁵⁵ Smoothing

⁵⁶ Mel Scale

⁵⁷ Log-Energy

⁵⁸ Frame

⁵⁹ Psychoacoustic

⁶⁰ Perceptual Linear Prediction

⁶¹ Cepstral

$$\Delta y_t^s = \frac{\sum_{i=1}^n w_i (y_{t+i}^s - y_{t-i}^s)}{2 \sum_{i=1}^n w_i^2}$$

که n طول پنجره است و w_i ها ضرایب رگرسیون می‌باشند. پارامترهای دلتا-دلتا، $\Delta^2 y_t^s$ ، به روش مشابهی ولی با استفاده از تفاضل‌های پارامترهای دلتا، بدست می‌آیند. وقتی اینها را به هم مرتبط کنیم، بردار ویژگی y_t بدست می‌آید:

$$y_t = [y_t^{sT} \quad \Delta y_t^{sT} \quad \Delta^2 y_t^{sT}]^T.$$

مدل‌های آکوستیکی HMM

همانطور که اشاره گردید، هر کلمه گفته‌شده w به K_w صدای پایه که آوای پایه⁶² نامیده می‌شوند شکسته می‌شود. این دنباله، تلفظ آن نامیده می‌شود:

$$q_{1:K_w}^{(w)} = q_1, \dots, q_{K_w}$$

برای اینکه امکان چندین تلفظ وجود داشته باشد، احتمال $P(Y|w)$ را می‌توان روی چند تلفظ حساب کرد:

$$p(Y|w) = \sum_Q p(Y|Q)P(Q|w),$$

که جمع روی همه دنباله‌های معتبر تلفظ w است و Q یک دنباله بخصوص از تلفظ‌ها است:

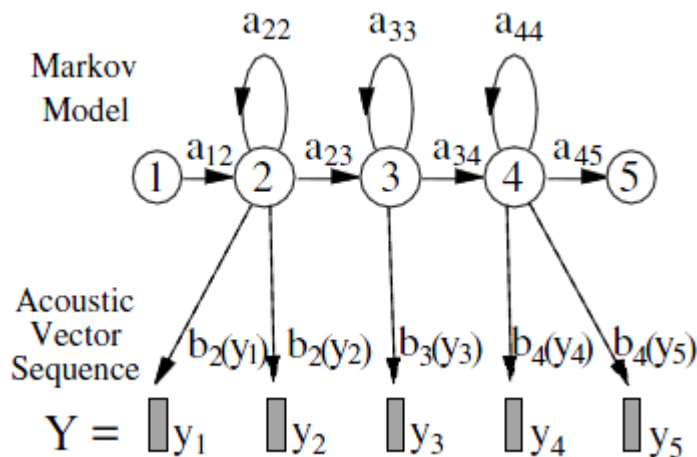
$$P(Q|w) = \prod_{l=1}^L P(q^{(w_l)}|w_l),$$

و هر $q^{(w_l)}$ یک تلفظ معتبر برای کلمه w_l است. در عمل، فقط تعداد کمی تلفظ مختلف برای هر w_l وجود دارد که حاصل جمع را راحتتر می‌کند.

هر آوای پایه q را با یک HMM پیوسته آنگونه که در شکل 2.2 آمده است نشان می‌دهیم و پارامترهای احتمال گذار را با $\{a_{ij}\}$ و توزیع‌های مشاهدات خروجی را با $\{b_j()\}$ نمایش

⁶² Base Phone

می‌دهیم. در عمل یک HMM از حالت کنونی خود به یکی از حالت‌های متصلش در هر مرحله زمانی گذار می‌کند.



شکل 2.2: مدل آوایی مبتنی بر HMM

احتمال یک گذار از حالت s_i به حالت s_j با احتمال گذار $\{a_{ij}\}$ نشان داده می‌شود. موقع ورود به یک حالت، یک بردار ویژگی با استفاده از توزیع متناسب به حالت ورودی، $\{b_j(\cdot)\}$ ، تولید می‌شود. این گونه فرایند به مفروضات استاندارد استقلال شرطی HMM منجر می‌شود:

- حالتها با فرض داشتن حالت قبلی، از همه حالت‌های دیگر مستقل شرطی هستند.
- مشاهدات با فرض داشتن حالتی که آن را تولید کرده است، از همه مشاهدات دیگر مستقل شرطی هستند.

برای توزیع خروجی از فرض گوسی چندمتغیره استفاده می‌شود:

$$b_j(y) = \mathcal{N}(y; \mu^{(j)}, \Sigma^{(j)}),$$

که $\mu^{(i)}$ میانگین حالت s_j و $\Sigma^{(i)}$ کوواریانس آن است. از آنجا که ابعاد بردار آکوستیکی y نسبتاً بالا است، غالباً کوواریانس‌ها قطری در نظر گرفته می‌شوند.

با فرض اینکه HMM Q ترکیبی با به هم پیوستن همه آواهای پایه $q^{(w1)}, \dots, q^{(wL)}$ شکل می‌گیرد، آنگاه احتمال آکوستیکی بدین طریق بدست می‌آید:

$$p(Y|Q) = \sum_{\theta} p(\theta, Y|Q),$$

که $\theta = \theta_0, \dots, \theta_{T+1}$ یک دنباله حالت در مدل ترکیبی است و

$$p(\theta, Y|Q) = a_{\theta_0\theta_1} \prod_{t=1}^T b_{\theta_t}(y_t) a_{\theta_t\theta_{t+1}}.$$

در این معادله، θ_0 و θ_{T+1} متناظر با حالت ورودی غیرقابل خروج و حالت خروج می‌باشند که در شکل 2.2 نشان داده شده است. اینها برای ساده‌سازی فرایند اتصال مدل‌های آوایی برای ساختن کلمات آورده شده‌اند. برای ساده‌سازی، در ادامه، از حالت‌های غیرقابل خروج صرف‌نظر کرده و تمرکز را روی دنباله حالت $\theta_1, \dots, \theta_T$ قرار می‌دهیم.

پارامترهای مدل آکوستیکی $\lambda = [\{a_{ij}\}, \{b_j()\}]$ را می‌توان به خوبی از روی پیکره‌ای که شامل گفتارهای آموزشی باشد با الگوریتم فرووارد-بکوارد بدست آورد که مثالی از بیشینه‌سازی امید ریاضی⁶³ یا EM می‌باشد. برای هر گفتار $Y^{(r)}$ ، $r=1, \dots, R$ با طول $T^{(r)}$ ، HMM‌هایی که متناظر با دنباله کلمات موجود در گفتار می‌باشند، پیدا شده و HMM ترکیبی مورد نظر ساخته می‌شود. در اولین مرحله الگوریتم یا E-step، احتمال فرووارد $\alpha_t^{(rj)} = p(Y_{1:t}^{(r)} | \theta_t = s_j; \lambda)$ و احتمال بکوارد $\beta_t^{(ri)} = p(Y_{t+1:T}^{(r)} | \theta_t = s_i; \lambda)$ با فرمول بازگشتی زیر محاسبه می‌شوند:

$$\alpha_t^{(rj)} = \left[\sum_i \alpha_{t-1}^{(ri)} a_{ij} \right] b_j(y_t^{(r)})$$

$$\beta_t^{(ri)} = \left[\sum_j a_{ij} b_j(y_{t+1}^{(r)}) \beta_{t+1}^{(rj)} \right],$$

⁶³ Expectation Maximisation

که i و j روی همه حالت‌ها جمع زده می‌شوند. وقتی این فرمول بازگشتی را محاسبه می‌کنیم، ممکن است برای قطعه‌های بزرگ گفتار under-flow رخ دهد، لذا در عمل از احتمالات لگاریتمی استفاده می‌شود.

با داشتن احتمالات فرورارد و بکوارد، احتمال اینکه مدل در زمان t در حالت s_j باشد برای هر گفتاری مثل r بدین صورت است:

$$\gamma_t^{(rj)} = P(\theta_t = s_j | Y^{(r)}; \lambda) = \frac{1}{P^{(r)}} \alpha_t^{(rj)} \beta_t^{(rj)},$$

که $P^{(r)} = p(Y^{(r)}; \lambda)$. این احتمالات حضور در حالت که تعداد تصرف⁶⁴ نیز نامیده می‌شوند، نمایانگر نوعی تنظیم نرم⁶⁵ حالت‌های مدل نسبت به داده‌ها می‌باشند و به راحتی می‌توان نشان داد که مجموعه جدید پارامترهای گوسی که توسط فرمول زیر تعریف می‌شوند، احتمال داده به شرط این تنظیمات را بیشینه می‌سازند:

$$\hat{\mu}^{(j)} = \frac{\sum_{r=1}^R \sum_{t=1}^{T^{(r)}} \gamma_t^{(rj)} y_t^{(r)}}{\sum_{r=1}^R \sum_{t=1}^{T^{(r)}} \gamma_t^{(rj)}}$$

$$\hat{\Sigma}^{(j)} = \frac{\sum_{r=1}^R \sum_{t=1}^{T^{(r)}} \gamma_t^{(rj)} (y_t^{(r)} - \hat{\mu}^{(j)})(y_t^{(r)} - \hat{\mu}^{(j)})^T}{\sum_{r=1}^R \sum_{t=1}^{T^{(r)}} \gamma_t^{(rj)}}$$

یک معادله باز-تقریب مشابه را می‌توان از احتمالات گذار استخراج نمود:

$$\hat{a}_{ij} = \frac{\sum_{r=1}^R \frac{1}{P^{(r)}} \sum_{t=1}^{T^{(r)}} \alpha_t^{(ri)} a_{ij} b_j(y_{t+1}^{(r)}) \beta_{t+1}^{(rj)} y_t^{(r)}}{\sum_{r=1}^R \sum_{t=1}^{T^{(r)}} \gamma_t^{(ri)}}.$$

این مرحله دوم الگوریتم یا M-step می‌باشد. با شروع از یک تقریب آغازین پارامترها، $\lambda^{(0)}$ ، تکرارهای متوالی الگوریتم EM منجر به مجموعه پارامترهای $\lambda^{(1)}, \lambda^{(2)}, \dots$ می‌شود که

⁶⁴ Occupation Counts

⁶⁵ Soft Alignment

احتمال را تا رسیدن به یک ماکزیموم محلی بهبود می‌بخشند. برای انتخاب مجموعه پارامتر اولیه $\lambda^{(0)}$ می‌توان از انتساب میانگین و کوواریانس کلی داده‌ها به توزیع‌های گوسی خروجی و احتمالات گذار یکسان استفاده نمود.

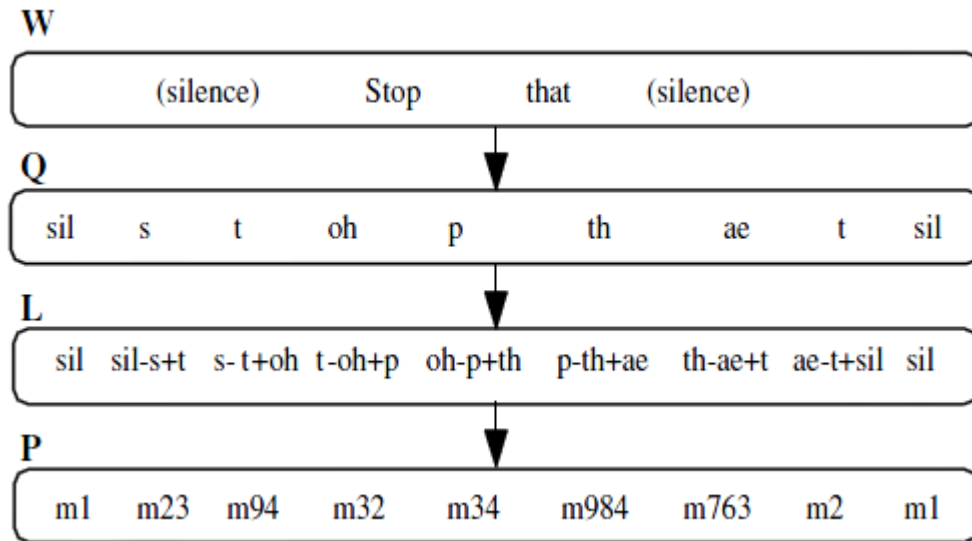
به این روش مدل‌سازی آکوستیکی، مدل دانه تسبیح⁶⁶ گفته می‌شود، چرا که همه اجزا گفتار با به هم پیوستن مدل‌های آوایی پشت سر هم مدل می‌شوند. مشکل اصلی این راهکار این است که شکستن هر کلمه واژگانی به دنباله‌ای از آواهای پایه مستقل از متن، تغییرات بسیار وابسته به متن موجود در گفتار واقعی را نادیده می‌انگارد. مثلاً تلفظ پایه کلمه mood و cool از یک مصوت یکسان oo بهره می‌گیرند ولی در عمل بازنمایی oo در دو بافت مختلف بدلیل تأثیر صامت‌های قبلی و بعدی، بسیار متفاوت است. مدل‌های آوایی مستقل از متن را تک‌آوا⁶⁷ می‌نامند.

یک راه ساده برای بهبود این مشکل، استفاده از یک مدل آوایی منحصر بفرد برای هر زوج همسایه چپ و راست می‌باشد. مدل‌های حاصله را مدل‌های سه‌آوا⁶⁸ می‌نامند و اگر N آوای پایه وجود داشته باشد، N^3 سه‌آوای بالقوه وجود دارد. برای پرهیز از مشکلات مربوطه به پراکندگی داده‌ها، مجموعه کامل سه‌آواهای منطقی L را می‌توان با خوشه‌بندی و دسته‌بندی پارامترهای هر خوشه، به یک مجموعه کاهش‌یافته از مدل‌های فیزیکی P نگاشت. این فرایند نگاشت در شکل 2.3 و دسته‌بندی پارامترها در شکل 2.4 نشان داده شده است که نماد $x-q+y$ نشانگر سه‌آوای متناظر به آوای گفته‌شده q در بافت آوای قبلی x و آوای بعدی y می‌باشد.

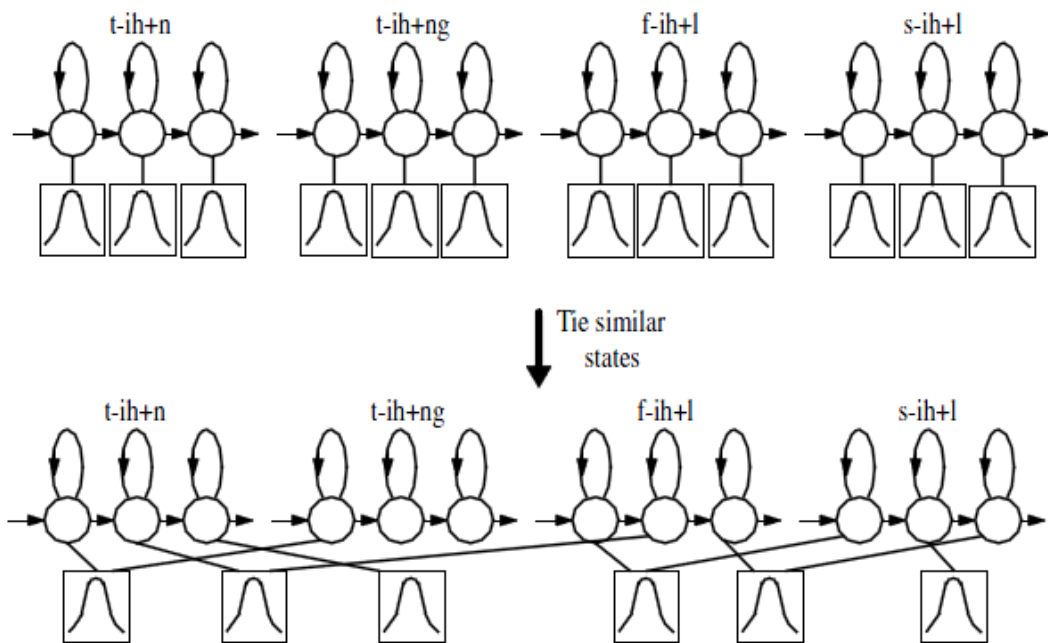
⁶⁶ Beads-on-a-string

⁶⁷ Monophone

⁶⁸ Triphone



شکل 2.3: مدل‌سازی آوای وابسته به متن



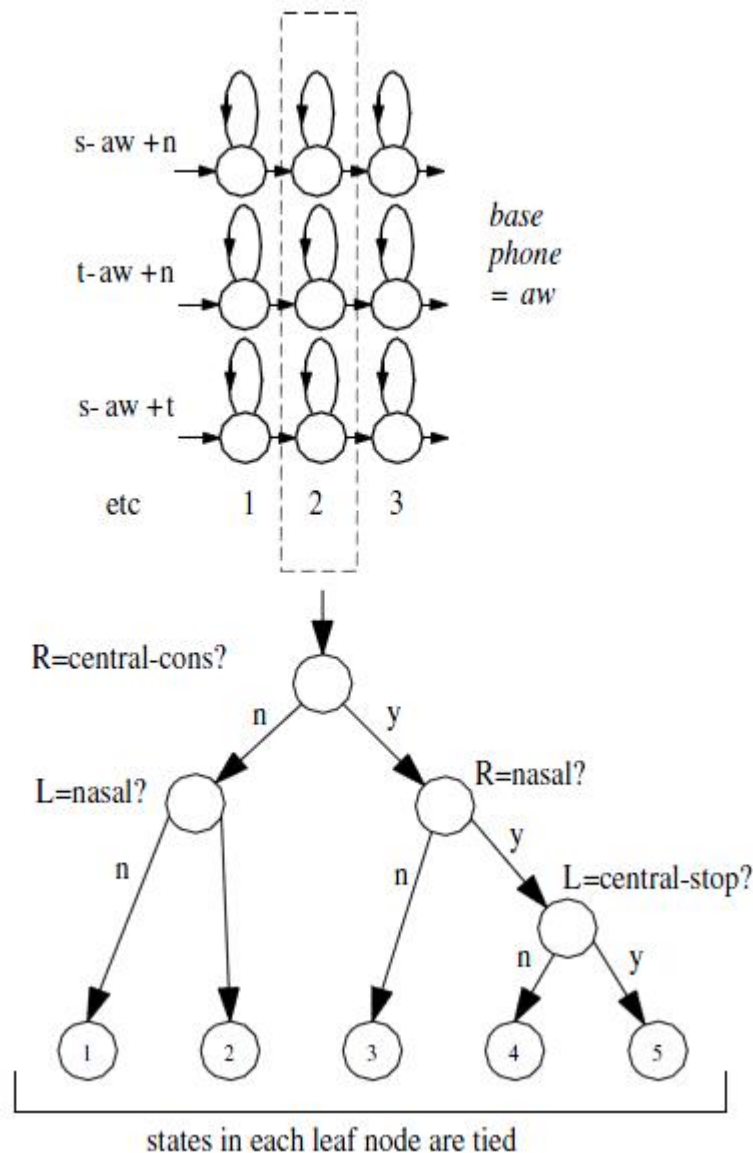
شکل 2.4: تشکیل مدل‌های آوا با حالت‌های دسته‌بندی‌شده

هر تلفظ پایه آوای q به سادگی با جستجو در یک لغتنامه تلفظی بدست می‌آید، سپس اینها براساس بافت به آوای منطقی نگاشته می‌شوند، و نهایتاً آواهای منطقی به آواهای فیزیکی

نگاشته می‌شوند. باید توجه داشت که وابستگی به متن بین مرز کلمات پخش می‌شود و این امر برای خیلی از فرایندهای مهم آوایی، یک مسأله اساسی می‌باشد. مثلاً /p/ در stop that انفجار⁶⁹ خود را به خاطر صامت بعدی از دست می‌دهد.

خوشه‌بندی مدل‌های منطقی به فیزیکی معمولاً در سطح-حالت عمل می‌کند نه در سطح-مدل، چرا که این حالت ساده‌تر است و اجازه می‌دهد تا مدل‌های فیزیکی گسترده‌تری با قدرت بالا تخمین زده شوند. انتخاب اینکه کدام حالت‌ها را باید دسته‌بندی کرد معمولاً با استفاده از درخت‌های تصمیم‌گیری انجام می‌شود. موقعیت حالت هر آوای q یک درخت دودویی متناظر دارد. هر گره این درخت حاوی سوالی درباره بافت است. برای خوشه‌بندی حالت i آوای q، همه حالت‌های i همه مدل‌های منطقی که از q استخراج شده‌اند را تحت تمرکز واحد در گره ریشه درخت درمی‌آوریم. بسته به پاسخ سوال هر گره، دسته حالتها متناوباً شکسته می‌شود تا جایی که همه حالت‌ها به گره‌های برگ برسند. سپس همه حالت‌های موجود در گره برگ دسته‌بندی می‌شوند تا یک مدل فیزیکی را شکل دهند. سوالات هر گره را از بین یک مجموعه از قبل معین انتخاب می‌کنیم تا احتمال داده آموزشی به شرط مجموعه نهایی دسته‌بندی حالت‌ها، بیشینه شود. اگر توزیع‌های خروجی حالت‌ها گوسی تک-مولفه باشند و تعداد تصرفات معلوم باشد، آنگاه افزایش احتمال حاصل از شکستن گوسی‌ها در هر گره را می‌توان به سادگی از روی تعدادها و پارامترهای مدل بدون ارجاع به داده آموزشی استخراج نمود. لذا درخت‌های تصمیم‌گیری با استفاده از یک الگوریتم شکستن گره تکراری حریصانه⁷⁰، می‌توانند به نحو موثری رشد کنند. شکل 2.5 این خوشه‌بندی مبتنی بر درخت را نشان می‌دهد. در این شکل، آوای منطقی s-aw+n و t-aw+n هر دو به گره برگ 3 اختصاص داده می‌شوند و لذا در یک حالت مرکزی یکسان مدل فیزیکی شریک خواهند بود.

⁶⁹ Burst⁷⁰ Greedy



شکل 2.5: خوشه‌بندی با استفاده از درخت تصمیم‌گیری

تقسیم‌بندی حالت‌ها با استفاده از درخت‌ها تصمیم‌گیری آوایی، مزیت‌های زیادی دارد. خصوصاً، مدل‌های منطقی‌ای که مورد نیاز هستند ولی اصلاً در داده آموزشی دیده نشده‌اند، به راحتی قابل تولید می‌باشند. یک عیب این روش این است که تقسیم‌بندی نسبتاً درشت⁷¹ است. این مشکل را می‌توان با استفاده از دسته‌بندی نرم⁷² تا حدی حل کرد. در این روش، یک مرحله پس-پردازشی وجود دارد که هر حالت را با یک یا دو همسایه نزدیکش گروه‌بندی می‌کند و همه

⁷¹ Coarse

⁷² Soft-Tying

گوسی‌های آنها را در یک دسته قرار می‌دهد. لذا مدل‌های گوسی ساده به مدل‌های گوسی ترکیبی ترکیب می‌شوند درحالی‌که تعداد کل گوسی‌ها در سیستم ثابت می‌ماند.

در کل، مدل‌های آکوستیکی اصلی یک سیستم مدرن بازشناسی گفتار از یک مجموعه از HMM‌های سه-حالتی دسته‌بندی‌شده تشکیل شده‌اند که توزیع‌های خروجی گوسی دارند. این سیستم طی مراحل زیر ساخته می‌شود:

1. یک مجموعه تک‌آوای flat-start ایجاد می‌شود که هر آوای پایه آن یک HMM تک-گوسی است و میانگین و کوواریانس آن با میانگین و کوواریانس داده آموزشی برابر است.
2. پارامترهای تک‌آوای گوسی با 3 یا 4 تکرار EM بدست می‌آیند.
3. هر تک‌آوای واحد گوسی q به ازای هر سه‌آوای $x-q+y$ که در داده آموزشی ظاهر می‌شود کپی می‌شود.
4. مجموعه حاصله سه‌آوای داده-آموزشی دوباره با استفاده از EM تقریب زده می‌شود و تعداد تصرف‌های حالت آخرین تکرار ذخیره می‌شوند.
5. برای هر حالت در هر آوای پایه یک درخت تصمیم‌گیری ساخته می‌شود، سه‌آوای داده-آموزشی به مجموعه کوچکتری از سه‌آوای دسته‌بندی‌شده نگاشته می‌شوند و با استفاده از EM دوباره تقریب زده می‌شوند. نتیجه نهایی، مجموعه مدل آکوستیکی وابسته به متن مورد نظر می‌باشد.

3. مدل‌های موضوع پنهان (مانند LSA، PLSA، LDA و...) کاربردهای مختلفی در پردازش زبان طبیعی دارند. در مورد این کاربردها و نحوه استفاده از این مدل‌ها در کاربردهای مختلف پردازش زبان طبیعی به تفصیل توضیح دهید.

گسترش مجموعه‌های عظیم مستندات الکترونیکی مانند وب، مقالات خبری، وبلاگ‌ها و مقالات علمی، چالش‌های جدید، متنوع و جالبی را پیش روی پژوهشگران در زمینه بازیابی اطلاعات و استخراج داده‌ها قرار داده است. علی‌الخصوص، نیاز روزافزونی به تکنیک‌های خودکار برای تحلیل و استخراج الگوهای جذاب از این مجموعه مستندات احساس می‌شود.

مدلسازی موضوع پنهان و یک تکنیک کاملاً بی‌سرپرست برای کشف موضوع در مجموعه‌های بزرگ مستندات اخیراً بسیار مورد توجه قرار گرفته است. این مدل‌ها مانند LDA و PLSA الگوهای باهم‌آیی⁷³ کلمات در مستندات را استخراج می‌کنند تا خوشه‌های احتمالی معنی‌دار آماری از کلمات را بدست آورند که موضوع نامیده می‌شود. این مدل‌ها همچنین یک احتمال عضویت به مستندات در فضای موضوع پنهان اختصاص می‌دهند تا بتوانیم مستندات را در یک فضای با ابعاد کمتر مشاهده و پردازش کنیم.

اکثر مدل‌های موجود در این چارچوب مانند مدل‌های موضوع پویا⁷⁴، تخصیص پاچینکو⁷⁵، مدل موضوع همبسته⁷⁶ و غیره، جنبه‌های گوناگونی از مجموعه مستندات مانند زمان، سلسله‌مراتب موضوعات و همبستگی بین موضوعات را به ترتیب مدل می‌کنند. هرچند همه مدل‌هایی که به آنها اشاره گردید یک ویژگی مهم موجود که حاوی اطلاعات ارزشمندی می‌باشد را نادیده می‌گیرند و آن ویژگی، ساختار ارجاع⁷⁷ یا هایپرلینک است. این یک حقیقت شناخته‌شده در زمینه بازیابی اطلاعات است که ارجاع بین دو مستند نه تنها نشانگر شباهت موضوعی آن دو مستند است بلکه نشانگر موثق بودن مستند مورد ارجاع است. این ایده با استفاده از الگوریتم‌هایی مثل PageRank که امروزه یک تکنیک رایج در تکنولوژی موتورهای جستجو می‌باشد بدست آمده است.

در یکی از اولین تلاش‌هایی که برای اعمال مدل‌های موضوع در مدل‌کردن داده‌های ارجاع انجام گردید، یک تعمیم از مدل PLSA ساخته شد که PHITS نامیده می‌شد. این مدل یک فرایند زیایا را نه تنها برای متن بلکه برای ارجاعات (هایپرلینک‌ها) تعریف می‌کند. ساختن هر هایپرلینک در مستند d بصورت یک نمونه‌برداری چندجمله‌ای از مستند هدف d' از روی توزیع خاص-موضوع Ω روی مستندات، مدل می‌شود. این مدل یک احتمال بالای Ω_{kd} به مستند d' با توجه به موضوع k نسبت می‌دهد، اگر که این مستند از چندین مستند دیگر که درباره این موضوع بحث می‌کنند لینک داده شده باشد. نشان داده شده است که بازنمایی مستند در فضای موضوع که از این مدل بدست آمده است، در مقایسه با بازنمایی بدست آمده از متن خالی، کارایی طبقه‌بندی‌کننده⁷⁸ مستند را افزایش و بهبود می‌بخشد. از این مدل به Link-PLSA یاد می‌شود.

⁷³ Co-Occurrence

⁷⁴ Dynamic Topic Model

⁷⁵ Pachinko Allocation

⁷⁶ Correlated Topic Model

⁷⁷ Citation

⁷⁸ Classifier

مدل مشابه دیگری که مدل عضویت آمیخته⁷⁹ نامیده می‌شود توسعه داده است که در آن PLSA با LDA بعنوان واحد سازنده اصلی تعویض شده است. از این مدل به Link-LDA یاد می‌شود. فرایند زایایی این مدل در جدول 3.1 نشان داده شده است و بازنمایی گرافیکی متناظر آن در شکل 3.1 آورده شده است. همانگونه که در شکل دیده می‌شود، فرایندهای تولید برای کلمات و هایپرلینک‌ها بسیار مشابه هستند و دارای یک توزیع موضوع خاص-مستند Ω یکسان برای تولید موضوعات پنهان خود می‌باشند. لذا این مدل (Link-PLSA) حاوی این نکته می‌باشند که مستنداتی که هایپرلینک‌های و کلمات مشترکی دارند، متمایلند که موضوع یکسانی داشته باشند.

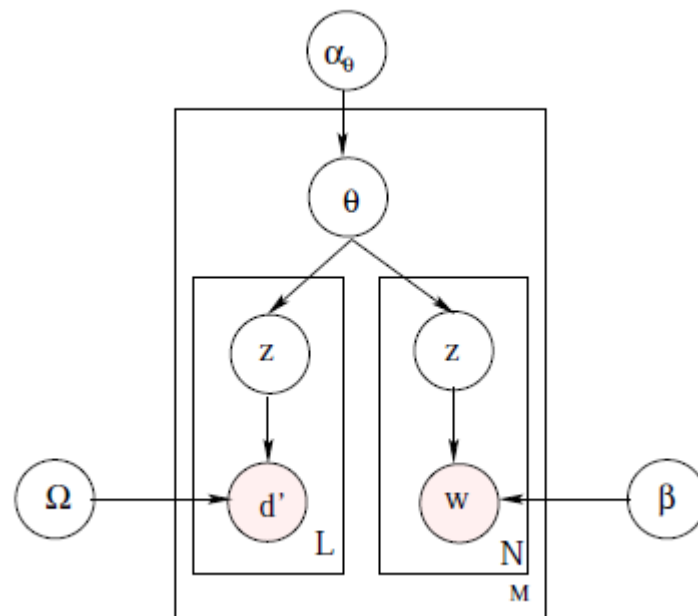
M	Total number of documents
M_{\leftarrow}	Number of cited documents
M_{\rightarrow}	Number of citing documents
V	Vocabulary size
K	Number of topics
N_{\leftarrow}	Total number of words in the cited set
d	A citing document
d'	A cited document
$\Delta(p)$	A simplex of dimension $(p - 1)$
$c(d, d')$	citation from d to d'
L_d	Number of hyperlinks in document d
N_d	Number of words in document d

جدول 3.1: نماد برخی متغیرهای زیاد بکاررفته

For each document $d = 1, \dots, M$
Generate $\theta_d \in \Delta(K) \sim \text{Dir}(\cdot \alpha_\theta)$
For each position $n = 1, \dots, N_d$
Generate $z_n \in \{1, \dots, K\} \sim \text{Mult}(\cdot \theta_d)$
Generate word $w_n \in \{1, \dots, V\} \sim \text{Mult}(\cdot \beta_{z_n})$
For each hyperlink $l = 1, \dots, L_d$
Generate $z_l \in \{1, \dots, K\} \sim \text{Mult}(\cdot \theta_d)$
Generate target doc. $d'_l \in \{1, \dots, M\} \sim \text{Mult}(\cdot \Omega_{z_l})$

جدول 3.2: فرایند تولید برای مدل Link-LDA

⁷⁹ Mixed Membership Model



شکل 3.1: بازنمایی گرافیکی مدل LDA

هر دوی Link-PLSA و Link-LDA هاپیرلینک را بصورت مقادیری که یک متغیر تصادفی اختیار می‌کند در نظر می‌گیرند. به عبارتی دیگر، این مدل‌ها خوشه‌های موضوعی احتمالی هاپیرلینک‌ها را دقیقاً به همان روشی که مدل‌های LDA و PLSA پایه خوشه‌های موضوعی کلمات را استخراج می‌کنند، بدست می‌آورند. این روش‌ها نمی‌توانند صریحاً روابط موضوعی بین متن مستند ارجاع‌دهنده و متن مورد ارجاع را مدل کنند. می‌توان امیدوار بود که کیفیت بهتری از موضوعات را با استخراج این اطلاعات اضافی بدست آورد.

اخیراً یک روش جدید مبتنی بر LDA ارائه شده است که جریان تأثیرگذاری از متن مستندات مورد ارجاع به متن مستندات ارجاع‌دهنده را امکان‌پذیر می‌سازد. در این راهکار، هر مستند ارجاع‌دهنده، در ساختن متن خود، از یکی از ارجاعاتش، موضوع به امانت می‌گیرد. در انتخاب ارجاعی که قرار است موضوع از آن به امانت گرفته شود، مستند از توزیع خودش روی ارجاعاتش استفاده می‌کند. این توزیع به عنوان تأثیر هر ارجاع روی مستند ارجاع‌دهنده تعبیر می‌شود. هرچند این مدل صریحاً موضوعیت ارجاعات را مدل نمی‌کند. بعلاوه این مدل ارجاعات را بعنوان داده ورودی در نظر می‌گیرد، درحالی‌که می‌توان مدل‌هایی را یافت که از روی مستندات مشاهده‌نشده، ارجاعات را تولید و پیش‌بینی می‌کنند.

Link-DLA دودو⁸⁰

در این مدل، مدل LDA با بلوک آماری عضویت آمیخته⁸¹ (MMSB) که قبلاً در مدل‌سازی تعاملات پروتئین-پروتئین استفاده شده بود، ترکیب می‌شود. مدل MMSB بر اساس تعاملات پروتئین‌ها، عضویت احتمالی به آنها برای موضوعات را بدینگونه نسبت می‌دهد: برای هر زوج پروتئین (d, d') ابتدا یک موضوع $z_{dd'}$ برای پروتئین d از توزیع θ_d آن روی موضوعات بدست می‌آوریم. بهمین طریق $z_{d'd}$ را از روی $\theta_{d'}$ بدست می‌آوریم. سپس حضور یا عدم‌حضور یک تعامل بین d و d' بصورت یک متغیر تصادفی دودویی از روی یک توزیع برنولی تولید می‌شود که پارامتر $\eta_{z_{dd'}, z_{d'd}}$ آن توسط موضوعات نمونه‌برداری شده از پروتئین‌های متناظر در این تعامل خاص مشخص می‌شود.

فرایند تولید

با در نظر گرفتن مستندات به مثابه پروتئین‌ها می‌توان این مدل را به متن نیز تعمیم داد. لذا برای هر زوج مستند، حضور یا عدم حضور یک ارجاع با یک متغیر تصادفی برنولی نمایش داده می‌شود. پارامتر این توزیع به موضوعات پنهانی که از هر کدام از این مستندات نمونه‌برداری شده‌اند، بستگی دارد. باید توجه داشت که تعامل پروتئین‌ها یک امر متقارن است. لیکن یک ارجاع، جهتدار است و ذاتاً نامتقارن می‌باشد. برای اینکه این مسأله را به حساب آوریم، برای هر زوج مستند (d, d') جهتداری ارجاع را بر اساس time-stampهای مستندات مشخص می‌کنیم. مثلاً اگر d' از d قدیمی‌تر باشد، آنگاه فرض می‌کنیم که ارجاع از d به سمت d' است. بعلاوه اگر $z_{dd'}$ موضوع پنهانی باشد که از روی d برای این تعامل نمونه‌برداری شده است و $z_{d'd}$ موضوع متناظر از d' برای همان تعامل باشد، آنگاه پارامتر برنولی بکاررفته برای تولید این ارجاع $\eta_{z_{dd'}, z_{d'd}}$ خواهد بود و نه $\eta_{z_{d'd}, z_{dd'}}$. عبارتی دیگر ماتریس η نامتقارن خواهد بود $(\eta_{kk'} \neq \eta_{k'k})$ تا جهتداری ارجاع در نظر گرفته شود.

فرایند تولید کلمات مانند فرایند LDA باقی خواهد ماند. باید توجه داشت که نسبت‌های موضوع خاص مستند θ_d که در تولید کلمات بکار می‌روند همانند آنی هستند که در تولید لینک‌های آن مستند بکار می‌روند. شرح مبسوطتری از فرایند تولیدی این مدل در جدول 3.3 آمده است. امکان ندارد که ارجاعات بین همه زوج مستندات را با یک نماد مسطح نمایش داد، لذا از یک بازنمایی گرافیکی که در شکل 3.2 آمده است استفاده می‌کنیم که در آن ارجاعات یک مجموعه از مستندات (مجموعه ارجاع‌دهنده) به یک مجموعه دیگر (مجموعه مورد ارجاع) را نشان

⁸¹ Mixed Membership Stochastic Block

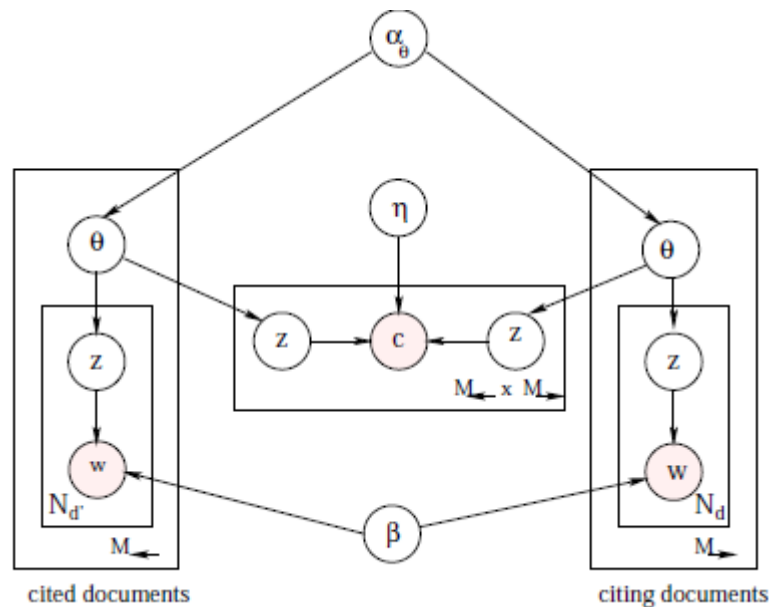
می‌دهیم. از روی شکل 3.2 معلوم است که موضوعیت مستند مورد ارجاع و مستند ارجاع‌دهنده صریحاً توسط ساختار V - در متغیر c که مشاهده شده است، وابسته شده‌اند.

```

For each document  $d = 1, \dots, M$ 
  Generate  $\theta_d \in \Delta(K) \sim \text{Dir}(\cdot | \alpha_\theta)$ 
  For each position  $n = 1, \dots, N_d$ 
    Generate  $z_n \in \{1, \dots, K\} \sim \text{Mult}(\cdot | \theta_d)$ 
    Generate  $w_n \in \{1, \dots, V\} \sim \text{Mult}(\cdot | \beta_{z_n})$ 
For each document pair  $(d, d')$ 
  Generate  $z_{dd'} \in \{1, \dots, K\} \sim \text{Mult}(\cdot | \theta_d)$ 
  Generate  $z_{d'd} \in \{1, \dots, K\} \sim \text{Mult}(\cdot | \theta_{d'})$ 
  Generate  $c_{d'd} \in \{0, 1\} \sim \text{Bernoulli}(\cdot | \eta_{z_{dd'} z_{d'd}})$ 

```

جدول 3.3: فرایند تولید برای LDA ارجاع دبدو



شکل 3.2: بازنمایی گرافیکی مدل LDA ارجاع دبدو

استنتاج واریاسیونی⁸²

شباهت لگاریتمی⁸³ داده مشاهده‌شده با توجه به این مدل بصورت زیر است:

⁸² Variational Inference

⁸³ Log-Likelihood

$$\begin{aligned}
& \log P(\mathbf{w}, \mathbf{c} | \alpha_\theta, \eta, \beta) \\
&= \log \left(\int_{\theta} \prod_{d=1}^M \left\{ P(\theta_d | \alpha_\theta) \prod_{n=1}^{N_d} \left(\sum_k \theta_{dk} \beta_{kw_n} \right) \right\} \right. \\
&\quad \times \prod_{d,d'} \left(\sum_{k,k'} \theta_{dk} \theta_{d'k'} (\eta_{k,k'})^{c_{dd'}} \right. \\
&\quad \left. \left. (1 - \eta_{k,k'})^{1-c_{dd'}} \right) d\theta \right)
\end{aligned}$$

از تقریب واریاسیونی mean-field برای توزیع خلفی⁸⁴ استفاده می‌کنیم:

$$\begin{aligned}
Q(\theta, \mathbf{z}) &= \prod_{d=1}^M \text{Dir}(\theta_d | \gamma_d) \left(\prod_{n=1}^{N_d} \prod_{k=1}^K \text{Mult}(z^{dnk} | \phi_{dn}) \right) \\
&\quad \times \prod_{d,d'} \prod_{k=1}^K \prod_{k'=1}^K \text{Mult}(z_{dd'k} | \lambda_{dd'}) \text{Mult}(z_{d' dk} | \lambda_{d' d'})
\end{aligned}$$

از جزئیات و مراحل استخراج استنتاج صرفنظر کرده و معادله نهایی را بدین شکل می‌نویسیم:

$$\phi_{dnk} \propto \beta_{kw_n} \exp(\Psi(\gamma_{dk})) \quad (1)$$

$$\begin{aligned}
\lambda_{dd'k} &\propto \exp(\Psi(\gamma_{dk}) + \sum_{k'} \lambda_{d' dk'} (c_{dd'} \log(\eta_{kk'}) \\
&\quad + (1 - c_{dd'}) \log(1 - \eta_{kk'}))) \quad (2)
\end{aligned}$$

$$\gamma_{dk} = \alpha_\theta + \sum_{n=1}^{N_d} \phi_{dnk} + \sum_{d'} \lambda_{dd'k} \quad (3)$$

$$\beta_{kv} \propto \sum_{d=1}^M \sum_{n=1}^{N_d} \phi_{dnk} \delta_v(w_n) \quad (4)$$

$$\eta_{kk'} = \frac{\sum_{d,d'} \lambda_{dd'k} \lambda_{d' dk'} c_{d,d'}}{\sum_{d,d'} \lambda_{dd'k} \lambda_{d' dk'}} \quad (5)$$

که $\Psi()$ تابع دیگاما⁸⁵ می‌باشد و $\delta_v(\mathbf{w})$ تابع دلتا می‌باشد با این فرض که $\delta_v(\mathbf{w})=1$ اگر $\mathbf{w}=\mathbf{v}$ و 0 در غیر اینصورت.

⁸⁴ Posterior Distribution

⁸⁵ Digamma

در موقع پیاده‌سازی، ابتدا، مرحله (1) را برای هر موقعیت در هر مستند اجرا می‌کنیم تا چندجمله‌ای واریاسیونی p محاسبه گردد. سپس برای هر زوج مستند (d, d') مرحله (2) را برای هر دو مستند آنقدر اجرا می‌کنیم تا زمانیکه پارامترهای $\lambda_{dd'}$ و $\lambda_{d'd}$ به هم نزدیک شوند. سپس با استفاده از مراحل (3) تا (5) و آمار محاسبه شده در مراحل (1) و (2)، مقادیر η و β را بروزرسانی می‌کنیم. این فرایند در یک حلقه خارجی و بیرونی آنقدر تکرار می‌شود تا زمانیکه حد پایین شباهت لگاریتمی کل داده آموزشی همگرا شود.

پیاده‌سازی استنتاج برای مستندات ارجاع‌دهنده کمی متفاوت است. مراحل (1)، (2) و (3) را مانند قبل انجام می‌دهیم ولی فقط مقادیر p و γ را برای مستندات ارجاع‌دهنده بروزرسانی کرده و γ را برای مستندات مورد ارجاع در مقادیری که موقع آموزش یادگرفته شد ثابت نگه می‌داریم. همچنین از مراحل (4) و (5) صرف‌نظر می‌کنیم چرا که فقط عمل استنتاج را داریم انجام می‌دهیم.

مدل Link-LDA دوبدو حقیقتاً یک مدل زایا برای متن و ارجاعات می‌باشد و قادر است هر ساختار لینک دلخواهی را مدل کند. با این حال، از آنجا که این مدل به مدل‌سازی صریح حضور یا عدم حضور لینک‌ها بین هر زوج مستند نیاز دارد، مدل‌سازی مجموعه‌های بزرگ مستندات بسیار پرهزینه خواهد بود. لذا یکی از مشکلات این مدل مقیاس‌پذیری آن است.

Link-PLSA-LDA

همانطور که ذکر شد، مدل Link-LDA دوبدو زیاد مقیاس‌پذیر نیست. مدل Link-LDA از طرفی دیگر، ارجاع را بصورت یک نمونه‌برداری چندجمله‌ای از متسند هدف مدل می‌کند و لذا به مقایسه همه زوج مستندات نیازی ندارد. در نتیجه مقیاس‌پذیرتر است. با این حال همانگونه که گفته شد، این مدل نمی‌تواند وابستگی موضوعی بیت مستندات ارجاع‌دهنده و مورد ارجاع را صریحاً مدل کند. با یک مصالحه، می‌توان مدل دیگری مثل Link-PLSA-LDA ارائه داد که بهترین خصوصیات این دو مدل را با هم ادغام کند. مدل جدید از فرایند نمونه‌برداری چندجمله‌ای برای تولید ارجاعات استفاده می‌کند و لذا مقیاس‌پذیری مدل Link-LDA را حفظ می‌کند. همزمان این مدل وابستگی موضوعی بین مستندات ارجاع‌دهنده و مورد ارجاع را صریحاً مدل می‌کند.

برای رسیدن به این هدف، مدل Link-PLSA-LDA جدید یک فرض ساده‌کننده می‌کند که ساختار لینک در پیکره یک گراف دو قسمتی⁸⁶ است و همه لینک‌ها از سمت مجموعه مستندات

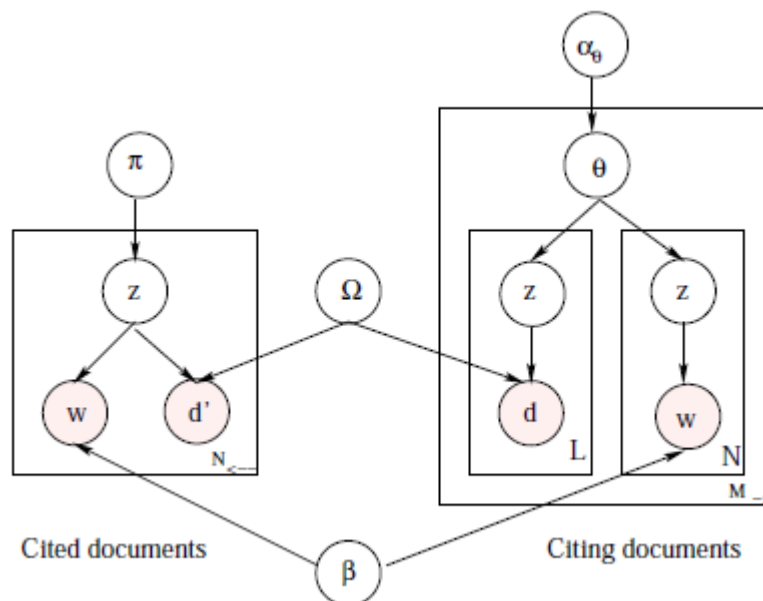
⁸⁶ Bipartite

ارجاع‌دهنده بیرون می‌آیند و به مجموعه مستندات مورد ارجاع اشاره می‌کنند. بعبارتی دیگر فرض می‌شود که هر مستند یا ارجاع‌دهنده است و یا مورد ارجاع و نه هر دو.

فرایند تولید

در این مدل، فرایند تولید محتوی و ارجاعات مستندات ارجاع‌دهنده مشابه روش Link-LDA است. بعلاوه به منظور مدل‌سازی صریح جریان اطلاعات از مستند ارجاع‌دهنده به مستند مورد ارجاع، یک فرایند تولید صریح برای محتوی مستندات مورد ارجاع تعریف می‌شود که از همان توزیع Ω استفاده می‌کند. در این فرایند تولیدی، مجموعه مستندات مورد ارجاع بصورت مکان‌هایی دیده می‌شوند که باید با کلمات پر شوند. ابتدا یک نسبت ادغام موضوع π برای کل مجموعه مستندات مورد ارجاع تهیه می‌شود. سپس کلمات به تعداد N بار درون مکان‌ها پر می‌شوند که N جمع کل طول مستندات مجموعه مستندات مورد ارجاع است، بدین ترتیب: هر بار یک موضوع k از نسبت ادغام π نمونه‌برداری می‌شود، سپس یک مکان d' از Ω_k برداشته و یک وقوع کلمه از β_k درون این مکان پر می‌شود. این فرایند دقیقاً مشابه پارامتربندی متقارن PLSA می‌باشد.

کل فرایند تولید با جزئیات مرحله به مرحله در جدول 3.4 و بازنمایی گرافیکی متناظر آن در شکل 3.3 آمده است. می‌توان مشاهده نمود که وابستگی‌های از مستندات مورد ارجاع به مستندات ارجاع‌دهنده طبق یک Ω مشاهده‌نشده منتشر می‌شود، مانند اصلی d-separation در شبکه‌های بی‌زی.



شکل 3.3: بازنمایی گرافیکی مدل Link-PLSA-LDA

Cited documents:

For $i = 1, \dots, N_{\leftarrow}$

Generate $z_i \in \{1, \dots, K\} \sim \text{Mult}(\cdot | \pi)$

Sample $d'_i \in \{1, \dots, M_{\leftarrow}\} \sim \text{Mult}(\cdot | \Omega_{z_i})$

Generate $w_i \in \{1, \dots, V\} \sim \text{Mult}(\cdot | \beta_{z_i})$

Citing documents:

For each citing document $d = 1, \dots, M_{\rightarrow}$

Generate $\theta_d \in \Delta(K) \sim \text{Dir}(\cdot | \alpha_{\theta})$

For each position $n = 1, \dots, N_d$

Generate $z_n \in \{1, \dots, K\} \sim \text{Mult}(\cdot | \theta_d)$

Generate $w_n \in \{1, \dots, V\} \sim \text{Mult}(\cdot | \beta_{z_n})$

For each citation position $l = 1, \dots, L_d$

Generate $z_l \in \{1, \dots, K\} \sim \text{Mult}(\cdot | \theta_d)$

Generate $d'_l \in \{1, \dots, M_{\leftarrow}\} \sim \text{Mult}(\cdot | \Omega_{z_l})$

جدول 3.4: فرایند تولیدی برای مدل Link-PLSA-LDA

احتمال داده مشاهده شده در این مدل به طریق زیر محاسبه می گردد:

$$\begin{aligned}
 & P(\mathbf{w}, \mathbf{c} | \pi, \alpha_{\theta}, \Omega, \beta) \\
 &= \prod_{n=1}^{N_{\leftarrow}} \left(\sum_k \pi_k \Omega_{k d'_n} \beta_{k w_n} \right) \\
 &\times \prod_{d=1}^{M_{\rightarrow}} \int_{\theta_d} (P(\theta_d | \alpha_{\theta}) \prod_{n=1}^{N_d} \left(\sum_k \theta_{dk} \beta_{k w_n} \right) \\
 &\times \prod_{l=1}^{L_d} \left(\sum_k \theta_{dk} \Omega_{k c_l} \right)) d\theta_d
 \end{aligned}$$

که \mathbf{w} کل متن مستندات ارجاع دهنده و مورد ارجاع است و \mathbf{c} مجموعه هایپرلینک ها/ارجاعات می باشد.

مانند مدل Link-LDA دودو، از تقریب واریاسیونی mean-field برای توزیع posterior متغیرهای مخفی استفاده می شود:

$$\begin{aligned}
Q(\theta, z, |w, c) &= \prod_{d=1}^{M_{\rightarrow}} (\text{Dir}(\theta_d | \gamma_d)) \\
&\times \prod_{n=1}^{N_d} \prod_{k=1}^K \text{Mult}(z_{dnk} | \phi_{dn}) \prod_{l=1}^{L_d} \prod_{k=1}^K \text{Mult}(z_{dlk} | \varphi_{dl}) \\
&\times \prod_{n=1}^{N_{\leftarrow}} \prod_{k=1}^K \text{Mult}(z_{d'nk} | \xi_{d'n})
\end{aligned}$$

با استفاده از رویه استاندارد استنتاج واریاسیونی، به معادله زیر می‌رسیم:

$$\phi_{dnk} \propto \beta_{kw_n} \exp(\Psi(\gamma_{dk})) \quad (6)$$

$$\varphi_{dlk} \propto \Omega_{kd'_l} \exp(\Psi(\gamma_{dk})) \quad (7)$$

$$\gamma_{dk} = \alpha_{\theta} + \sum_{n=1}^{N_d} \phi_{dnk} + \sum_{l=1}^{L_d} \varphi_{dlk} \quad (8)$$

$$\xi_{d'nk} \propto \Omega_{kd'} \beta_{kw_n} \pi_k \quad (9)$$

$$\beta_{kv} \propto \sum_{d'=1}^{M_{\leftarrow}} \sum_{n=1}^{N_{d'}} \xi_{d'nk} \delta_v(w_n) + \sum_{d=1}^{M_{\rightarrow}} \sum_{n=1}^{N_d} \phi_{dnk} \delta_v(w_n) \quad (10)$$

$$\pi_k \propto \sum_{d'=1}^{M_{\leftarrow}} \sum_{n=1}^{N_{d'}} \xi_{d'nk} \quad (11)$$

$$\Omega_{kd'} \propto \sum_{n=1}^{N_{d'}} \xi_{d'nk} + \sum_{d=1}^{M_{\rightarrow}} \sum_{l=1}^{L_d} \varphi_{dlk} \delta_{d'}(d'_l) \quad (12)$$

این بروزرسانی‌ها مرتباً تکرار می‌شوند تا به همگرایی برسند. از آنجاییکه بروزرسانی‌های مراحل (6) تا (8) به یکدیگر وابسته هستند، باید یک حلقه تکرار داخلی هم درون این معادلات انجام شود تا همگرا شوند.

برای استنتاج فقط روی مستندات ارجاع‌دهنده، فقط بین مراحل (6) تا (8) تا رسیدن به همگرایی، تکرار انجام می‌شود.

از این روش می‌توان روی داده‌های لینک‌شده مثل Citeseer یا داده‌های وبلاگ‌هایی که شامل هایپرلینک هستند استفاده نمود.

4. Natural Language Understanding (NLU) و Spoken (SLU) Language Understanding به روش‌های مختلفی انجام می‌گیرد که بعضی آماری، بعضی قاعده‌مند و بعضی ترکیبی هستند. روش‌های مختلف NLU و SLU را بررسی کنید و این روش‌ها را با جزئیات شرح دهید.

دو راهکار اصلی برای پردازش معنایی در فهم زبان گفته‌شده وجود دارد: یک راهکار مبتنی بر قاعده و یک راهکار آماری. می‌توان این دو روش را با هم ترکیب کرد و از شبکه‌های بیزی پویا⁸⁷ (DBNs) آماری و نحوی همراه با مدل‌های گرافی⁸⁸ (GMs) استفاده نمود. GMها با یک روش پیچیده ریاضی، احتمالات را با نظریه گراف ادغام می‌کنند. این امر به چهار تنظیمات⁸⁹ متفاوت منجر می‌شود که پیچیدگی آنها افزایش می‌یابد.

پردازش معنایی یکی از کلیدی‌ترین مولفه‌های سیستم‌های دیالوگ گفتاری⁹⁰ است. آن پرس‌وجوی⁹¹ کاربر را تحلیل کرده و به مدیر دیالوگ اجازه می‌دهد تا تصمیم‌های حساس به بافت درباره دیالوگ بگیرد. هر جمله با مشخص کردن مفاهیم دیکود می‌شود، مفاهیمی مانند “مبدأ”، “مقصد”، یا “زمان” و طبقات کلمات به هر مفهوم تعلق دارند، طبقاتی مانند “شهرها”، “از-مکان”⁹²، یا “تا-مکان”⁹³. یک مفهوم به کلمات وابسته است، هرچند فقط کلمات خاصی ممکن است به یک مفهوم تعلق داشته باشند. در نتیجه، نام یک شهر فقط بعنوان یک “شهر” مشخص نمی‌شود، بلکه بعنوان مفهومی هم است که کلمات درون آن رخ می‌دهند. بهترین فرضیه مفهوم کلمه با یک روش بیشترین شباهت⁹⁴ (ML) و الگوریتم معروف ویتربی بدست می‌آید. با این حتم قواعد گرامری باید صریحاً تعریف شوند. می‌توان از راهکار دیکود سلسله‌مراتبی استفاده کرد با این مزین که الزام‌های گرامری⁹⁵ با مدل‌های گرافی مدل می‌شوند. پارامترهای آنها با یک مجموعه آموزشی برچسب خورده⁹⁶، بطور خودکار یاد گرفته می‌شوند.

⁸⁷ Dynamic Bayesian Networks

⁸⁸ Graphical Models

⁸⁹ Setup

⁹⁰ Spoken Dialog Systems

⁹¹ Query

⁹² Fromloc

⁹³ Toloc

⁹⁴ Maximum Likelihood

⁹⁵ Grammatical Bindings

⁹⁶ Annotated

پردازش معنایی بصورت نگاشت خودکار بین کلمات W که خروجی بازشناسی گفتار می‌باشند به دنباله‌ای از طبقات کلمات L و مفاهیم C که برای انجام عمل فهم لازم هستند، تعریف می‌شود. مثالی از سیستم اطلاعاتی پرواز هوایی (ATIS) را می‌توان در شکل 4.1 مشاهده نمود که مفاهیم، معنایی هستند و برچسب‌ها طبقات کلمات می‌باشند. تمام مخفف‌های بکاررفته در ATIS شرح داده خواهند شد.

Sentence: Flight four sixteen departs Dallas at 9 : 10 A.M. Correct ?
Concepts: FN FN FN OR OR TD TD TD TD TD DU DU
Labeling : IN NU NW2 FR CI AD N10 IT N60 IT DU DU

شکل 4.1: طبقه‌بندی بصورت طبقات کلمات و مفاهیم

می‌توان از گرامرهای مستقل از متن بسطیافته⁹⁷ (ECFG) برای ساختن شبکه گذار وزندار استفاده نمود. ECFGها که ساختار زبان را می‌سازند، توسط یک متخصص بصورت دستی ساخته می‌شود. هدف اصلی، یادگیری خودکار ساختار معنایی زبان از روی یک پیکره و استفاده از قواعد گرامری اضافی (مانند ECFGها) برای قسمت‌هایی از زبان گفته‌شده است. بنابراین یک بازنمایی گرافیکی که اجازه مدل‌سازی سریع را می‌دهد، آموزش و دیکود EM همراه با الگوریتم ویتربی مناسب‌ترین راه است.

مدل شبکه بیزی پویا

مدل‌های گرافی (GMS)، ترکیبی از احتمالات و نظریه گراف هستند و یک زبان گرافیکی تصویری و الگوریتم‌های کارا و موثری برای محاسبه احتمالات و تصمیم‌گیری فراهم می‌سازند. یک شبکه بیزی (BN) نوعی از GM است که در آن گراف‌ها جهتدار و غیرحلقه‌ای⁹⁸ هستند. توزیع توأم احتمال (که ویژگی فاکتوربندی جهتدار⁹⁹ نامیده می‌شود) روی X به صورت زیر فاکتوربندی می‌شود:

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i | \text{parents}(x_i))$$

شبکه‌های بیزی پویا (DBNs) تعمیمی از BNها هستند و از آنها برای توصیف سری‌های زمانی استفاده می‌شود: یک BN یک برش از زمان را نشان می‌دهد. علاوه بر این یال‌های خط‌تیره‌دار وابستگی‌های موجود بین برش‌های زمانی متعاقب را توصیف می‌کنند. برای یک مشاهده O با طول T، DBN باز می‌شود: برش‌های زمانی به تعداد (T-2) باز تکرار می‌شوند

⁹⁷ Extended-Context-Free Grammars

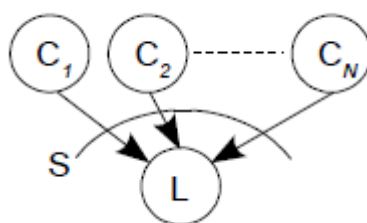
⁹⁸ Acyclic

⁹⁹ Directed Factorization Property

و از طریق یال‌های درونی خود به هم متصل می‌شوند. از آنها می‌توان برای فهم زبان استفاده نمود. علاوه بر این می‌توان از “والدهای راهگزینی”¹⁰⁰ برای اجماع قواعد نحوی و راهکارهای آماری استفاده نمود.

والدهای راهگزینی

معمولاً یک متغیر فقط یک مجموعه والد دارد. در شکل 4.2 متغیر S یک مفهوم را از بین $\{C_1, \dots, C_N\}$ انتخاب می‌کند، لذا، آن مفهوم تنها والد برجسب‌های L است.



شکل 4.2: والدهای راهگزینی در مدل‌های گرافی

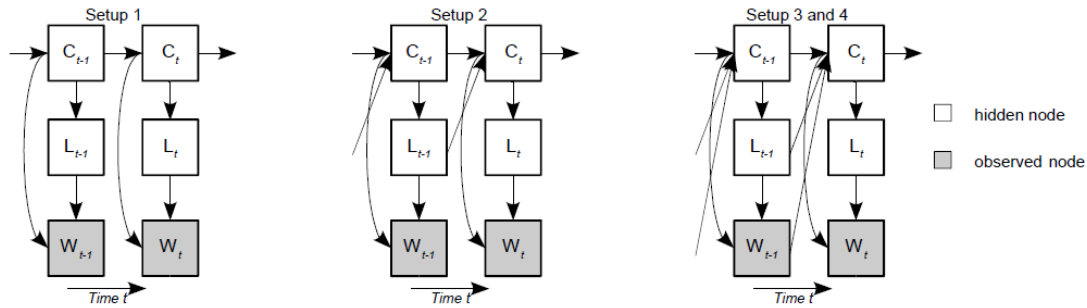
می‌توان از والدهای راهگزینی برای استفاده از راهکار مبتنی بر قواعد گرامری استفاده نمود. بعلاوه در پیکره‌های امروزی و جدید همه ترکیبات ممکن مثلاً در “زمان”، “روز هفته”، “تاریخ”، و “نمادهای عددی” در دسترس نیستند. این مفاهیم را می‌توان با والدهای راهگزینی پیاده‌سازی کرد.

مدل‌های فهم زبان

GM ای که برای مدل‌کردن تفسیر معنای زبانی بکار می‌بریم از سه گره در هر برش زمانی تشکیل می‌شود. لذا مسأله با سه متغیر برای هر کلمه جمله مدل می‌شود. از چهار تنظیمات متفاوت برای مدل‌کردن روابط مختلف بین این متغیرها استفاده می‌شود. لذا درحایکه گره‌های مدل در همه سه GM یکسان باقی می‌مانند، کمانه‌ای بین آنها متفاوت است. بنابراین هر GM یک فاکتوربندی مختلف از مسأله را توصیف می‌کند. این مراحل را باید با یک سیستم baseline مقایسه کرد که در آن بیشترین شباهت مفاهیم و برجسب‌های کلمات دیکود شده‌اند. برای استفاده از مفاهیم و برجسب‌ها، با تنظیمات 1 شروع می‌کنیم (شکل 4.3). باید توجه داشت که این تنظیمات و GM‌های بعدی نمایانگر مدل‌های مارکف مخفی فاکتوریل¹⁰¹ هستند.

¹⁰⁰ Switching Parents

¹⁰¹ Factorial Hidden Markov Models



شکل 4.3: تنظیمات مدل گرافی استفاده شده

در تنظیمات 1 وابستگی مفاهیم مدل می‌شود. با مشاهده اینکه مفاهیم فقط به مفاهیم قبلی وابسته نیستند بلکه به برجسب‌های قبلی هم وابسته هستند

(مثلاً *from-fromloc-origin, Dallas-city-origin*)، این مدل به تنظیمات 2 بسط داده می‌شود. در تنظیمات 3 وابستگی به کلمات قبلاً مشاهده شده بررسی می‌شود و نهایتاً در تنظیمات 4، قواعد گرامری توسط مفهوم GM ای “والدهای راه‌گزین” بکار گرفته می‌شوند.

حال احتمالات توأم چهار تنظیمات را شرح می‌دهیم. فرض می‌کنیم پیکره‌ای با N_W کلمه متفاوت W داریم، کلماتی که معنی یکسانی دارند در N_L طبقه کلمه (برجسب‌ها) گروه‌بندی می‌شوند. می‌توان بیشتر از یک طبقه کلمه را به N_C مفهوم مختلف C نگاشت. در ادامه گره‌های مدل‌ها و احتمالات زیرین آنها شرح داده خواهند شد که x, y, z نشانگر مشاهدات هستند:

$$P(C = \vec{x}) = \sum_{i=1}^{N_C} c_i \cdot \delta(\vec{x} - \vec{\mu}_i) \quad \text{with} \quad \sum_{i=1}^{N_C} c_i = 1,$$

$$P(L = \vec{y}) = \sum_{j=1}^{N_L} l_j \cdot \delta(\vec{y} - \vec{\nu}_j) \quad \text{with} \quad \sum_{j=1}^{N_L} l_j = 1,$$

$$P(W = \vec{z}) = \sum_{j=1}^{N_W} w_j \cdot \delta(\vec{z} - \vec{\omega}_j) \quad \text{with} \quad \sum_{j=1}^{N_W} w_j = 1.$$

تنظیمات 1

یک نمودار تصویری از GM در شکل 4.3 نشان داده شده است. هر ستون معرف یک برش زمانی است. گره بالایی C_t مفهوم زیرین کلمه جاری را مدل می‌کند. این مفهوم مشاهده نشده است و لذا بصورت مخفی مدل می‌شود. در اولین برش زمانی $t = 1$ ، مفهوم به هیچ متغیر

دیگری وابسته نیست و با احتمال $p(c_1) = 1$ نشان داده می‌شود، که توزیع اولیه مفهوم است. در همه برش‌های زمانی بعدی، مفهوم فقط به مفهوم برش‌های زمانی قبلی وابسته است: $p(c_t|c_{t-1})$. 1. لذا در این GM دنباله مفاهیم با یک زنجیره مارکف مرتبه اول نشان داده شده است. گره هر برچسب جستجو شده L_t در همه برش‌های زمانی فقط روی مفهوم C_t کلمه جاری W_t مشروط شده است و مشاهده نشده است و لذا مخفی است. از آنجاییکه هیچ تعاملی بین برچسب‌ها در برش‌های زمانی متوالی وجود ندارد، این مدل یک برچسبی را نشان می‌دهد که مستقل از برچسب قبلی رسم شده است ولی به مفهوم جاری C_t با احتمال $p(L_t|C_t)$ وابسته است. در نهایت کلمات دنباله مشخص می‌شوند و بصورت گره‌های مشاهده شده W_t مدل می‌شوند. هر کلمه جاری به برچسب و مفهوم برش زمانی جاری وابسته است. باز هم کلمات متوالی تعامل مستقیمی ندارند، ولی از طریق زنجیره مارکف مفهوم به هم متصلند. یعنی با دانستن مفهوم C_t کلمات متوالی W_t و W_{t-1} مستقل می‌باشند. در این مدل، احتمال یک کلمه بصورت $p(W_t|L_t, C_t)$ نمایش داده می‌شود که مشابه یک HMM فاکتوریلی است. رویهم‌رفته، GM تنظیمات 1 احتمال توأم دنباله کلمات، برچسب‌ها و مفاهیم را بدین صورت فاکتوربندی می‌کند:

$$p(V_1) = p(C_1) \cdot p(L_1|C_1) \cdot p(W_1|C_1, L_1) \cdot \prod_{t=2}^T p(C_t|C_{t-1}) \cdot p(L_t|C_t) \cdot p(W_t|C_t, L_t).$$

تنظیمات 2

تنظیمات 2، علاوه بر تنظیمات 1، شامل وابستگی به برچسب‌های قبلی L_{t-1} نیز می‌باشد. مثلاً هر دو برچسب “به-مکان” و “شهر” به مفهوم “مقصد” تعلق دارند. با داشتن برچسب قبلی، مثلاً “به-مکان”، تعیین برچسب بعدی در مفهوم “مقصد” مثلاً “شهر” مستحکم‌تر می‌باشد. در این حالت خاص، مشخص می‌شود که برچسب “شهر” به مفهوم “مقصد” تعلق دارد. گره بالایی C_t مفهوم زیرین کلمه جاری را مدل می‌کند. برخلاف تنظیمات 1، بعد از برش زمانی اول، توزیع مفهوم بر روی برچسب و مفهوم برش زمانی قبلی مشروط می‌شود: $p(C_t|C_{t-1}, L_{t-1})$. در این GM، مفهوم C_t از دانش برچسب‌های قبلی C_t و L_{t-1} سود می‌برد. باقی گره‌ها، شرایط و احتمالات متناظر مانند تنظیمات 1 می‌باشد. در کل GM تنظیمات 2 احتمال توأم دنباله کلمات، برچسب‌ها و مفاهیم را بدین صورت فاکتوربندی می‌کند:

$$p(V_2) = p(C_1) \cdot p(L_1|C_1) \cdot p(W_1|C_1, L_1) \cdot \prod_{t=2}^T p(C_t|C_{t-1}, L_{t-1}) \cdot p(L_t|C_t) \cdot p(W_t|C_t, L_t).$$

تنظیمات 3

بعنوان تعمیمی از تنظیمات 2، در این تنظیمات نه تنها برچسب قبلی بلکه کلمه قبلی نیز در تعیین مفهوم مشارکت دارند. مثلاً کلمات موجود در یک شماره پرواز مثل “چهار”، “شصت-پنج” و “صد” به برچسب‌های خاصی طبقه‌بندی می‌شوند (number word 1, “number 10”, “number 60”, “number rest”). با داشتن کلمات و برچسب‌های قبلی، تعیین برچسب بعدی در مفهوم مثلاً “شماره پرواز” در مقایسه با تنظیمات 2 بهبود می‌یابد. در این GM، مفهوم C_t از دانش برچسب‌های قبلی C_{t-1} , L_{t-1} , W_{t-1} سود می‌برد که با احتمال گذار $p(C_t|C_{t-1}, L_{t-1}, W_{t-1})$ بیان می‌شود. رویهم‌رفته GM در تنظیمات 3 احتمال توأم دنباله کلمات، برچسب‌ها و مفاهیم را بدین صورت فاکتوربندی می‌کند:

$$p(V_3) = p(C_1) \cdot p(L_1|C_1) \cdot p(W_1|C_1, L_1) \cdot \prod_{t=2}^T p(C_t|C_{t-1}, L_{t-1}, W_{t-1}) \cdot p(L_t|C_t) \cdot p(W_t|C_t, L_t).$$

تنظیمات 4

این تنظیمات از لحاظ گرافی قابل تمییز از تنظیمات 3 نیست، بنابراین همان فاکتوربندی‌های احتمال ترکیبی استفاده می‌شوند. هرچند مسأله کلمات ناموجود در پیکره مثل تاریخ‌ها، زمان‌ها و اعداد خاص با استفاده از قواعد گرامری در این تنظیمات حل می‌شوند. مثلاً شهر موجود در مفهوم “مبدأ” اغلب از طبقه کلمه “از-مکان” بدست می‌آید. بنابراین معادله زیر با استفاده از والد‌های راهگزين پیاده‌سازی می‌شود:

$$p(C_t = \text{“origin”} | L_{t-1} = \text{“fromloc”}) = 1$$

پارامترها و طبقه‌بندی‌ها¹⁰²

هر معنی معناشناختی (M) در گفتارهای زبان طبیعی را می‌توان با تنظیمات GM $x = \{1, 2, 3, 4\}$ و با پارامترهای $M_x = \{W_x, L_x, C_x\}$ توصیف نمود.

¹⁰² Classifications

پارامترهای مدل M_x برای هر N_C طبقه مفاهیم و N_L طبقه کلمات با الگوریتم EM در مرحله آموزش یادگرفته می‌شوند. طی طبقه‌بندی یک گفتار نامعلوم با کلمات W ، پارامترهای M_x را می‌توان برای هر مدل $P(L_x, C_x|W)$ با بیشترین شباهت محاسبه کرد:

$$P_x = \underset{L_x, C_x}{\operatorname{argmax}} P(C_x, L_x | W_T)$$

$$= \underset{L, C}{\operatorname{argmax}} \frac{P(V_x)}{P(W_T)} \quad \text{with } P(W_T) = \prod_{t=1}^T p(w_t).$$

اعمال الگوریتم ویتربی روی همه گفتارهای طبیعی در تنظیمات مختلف، به قطع‌بندی¹⁰³ معنایی متفاوتی از آنها می‌انجامد.

یکی از راهکارهای دیگر استفاده از بازنمایی شکاف-مقدار¹⁰⁴ است که در سیستم‌های دیالوگ بسیار کاربرد دارد. جدول 4.1 مثالی از شکاف-مقدارها را بعنوان یک بازنمایی معنایی نشان می‌دهد.

USER: what flights are there arriving in Chicago on continental airlines after 11pm
GOAL = FLIGHT
TOLOC.CITY_NAME = Chicago
AIRLINE_NAME = continental airlines
ARRIVE.TIME.TIME_RELATIVE = after
ARRIVE.TIME.TIME = 11pm

جدول 4.1: شکاف-مقدارها بعنوان بازنمایی معنایی

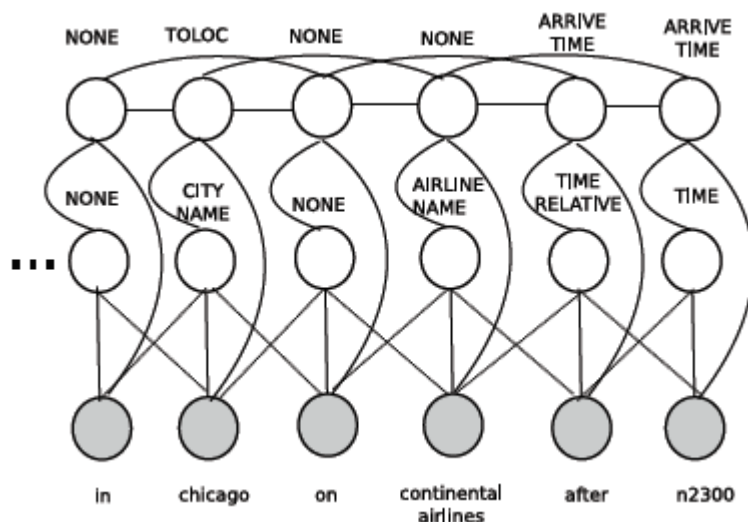
یکی از مدل‌های آماری برای SLU شبکه منطقی مارکف¹⁰⁵ (MLN) می‌باشد. یک MLN مجموعه‌ای از فرمول‌های وزن‌دار منطق مرتبه اول (FOL) می‌باشد که بعنوان الگو و قالبی برای مقداردهی اولیه شبکه‌های مارکف (MNs) بکار می‌رود. MLN‌ها مخصوصاً از نظر مدل‌سازی زبانی قابل توجه می‌باشند چرا که به راحتی با ویژگی‌های جدید قابل گسترش هستند و اجازه استفاده از روابط پیچیده بین گره‌های شبکه را می‌دهند. شکل 4.4 یک MN را برای بازنمایی شکاف-مقدار نشان می‌دهد. در این حالت، گره‌های روشن‌تر نمایانگر متغیرهای مخفی

¹⁰³ Segmentation

¹⁰⁴ Slot-Value

¹⁰⁵ Markov Logic Network

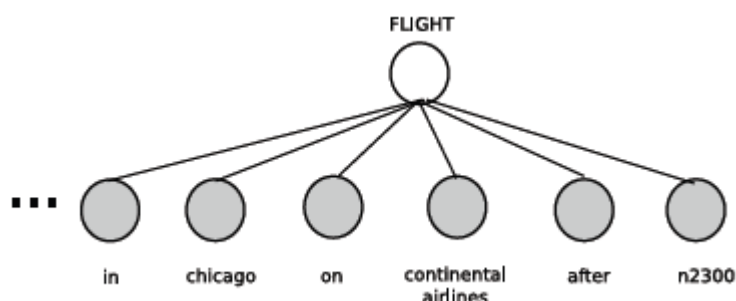
است (شکافهایی که قرار است تولید شوند) و گره‌های تیره‌تر نمایانگر متغیرهای قابل مشاهده می‌باشند (کلمات گفتار).



شکل 4.4: شبکه مارکف بصورت شکاف-مقدارها

مدل MLN

کار SLU را به دو قسمت می‌شکنیم. اولین کار شامل مدل کردن مولفه GOAL شکاف-مقدارها می‌باشد. شکل 4.5 یک MN را برای هدف مثال ما نشان می‌دهد. می‌توان توجه نمود که مولفه GOAL به کل گفتار وابسته است. کار دوم شامل مدل کردن بقیه شکاف‌ها می‌باشد. شکل 4.4 یک MN دولایه را برای شکاف‌های مثال ما نشان می‌دهد. لایه اول را می‌توان به صورت یک named entity دید، درحالی‌که لایه دوم یک تغییردهنده/تابع¹⁰⁶ برای named entityها می‌باشد.



شکل 4.5: شبکه مارکف برای شکاف هدف

با این مدل دولایه می‌توان روابط بین مولفه‌های تشکیل‌دهنده شکاف‌ها را بدست آورد. این امر با مشخص کردن یال‌های بین دولایه با استفاده از FOL-MLN حاصل می‌شود. این مدل

¹⁰⁶ Modifier/Function

همچنین مفروضات مارکف مرتبه اول و مارکف مرتبه دوم را برای لایه دوم فرض می‌کند. با این کار می‌توان هرگونه وابستگی درون دنباله شکاف‌ها را بدست آورد.

استخراج ویژگی

در MLN ها این امکان وجود دارد که متغیرهای قابل مشاهده بیشتری را اضافه نمود که با کلمات ورودی در ارتباط باشند. برای این منظور، می‌توان از منابع موجود استفاده کرد: الف) POS tag های کلمات گفتار با استفاده از TnT tagger و ب) قطعه‌های¹⁰⁷ نحوی کلمات گفتار با استفاده از CASS chunker. با این اطلاعات ویژگی زیر را برای مدل شکاف تعریف می‌کنیم: املا و POS tag کلمه برای پنجره‌ای از دو کلمه قبلی و بعدی؛ یک ویژگی دودویی که آیا کلمه، عدد است یا نامعلوم است؛ و کلمات هسته قطعه‌های نحوی. می‌توان نشان داد که با استفاده از یک مدل دولایه و الحاق ویژگی‌های اضافی می‌توان کارایی SLU را بهبود بخشید.

¹⁰⁷ Chunk